

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:39:50 ; Search time 10.6875 Seconds
(without alignments)
35.630 Million cell updates/sec

Title: US-09-228-866-1
Perfect score: 54
Sequence: 1 CNSRLHLRC 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	54	100.0	9	1	US-08-526-710-1
2	54	100.0	9	3	US-08-862-855-1
3	54	100.0	9	3	US-09-226-985-1
4	54	100.0	9	4	US-09-227-906-1
5	46	85.2	9	1	US-08-526-710-5
6	46	85.2	9	3	US-08-862-855-5
7	46	85.2	9	3	US-09-226-985-5
8	46	85.2	9	4	US-09-227-906-5
9	36	66.7	781	4	US-09-738-946-8
10	35	64.8	397	4	US-09-252-991A-28422
11	35	64.8	511	4	US-09-252-991A-28223

12	35	64.8	573	3	US-09-330-740A-8	Sequence 8, Appli
13	35	64.8	1402	4	US-09-004-838-11	Sequence 11, Appl
14	34	63.0	36	4	US-09-082-358B-88	Sequence 88, Appl
15	34	63.0	40	4	US-09-227-357-580	Sequence 580, App
16	34	63.0	50	3	US-08-685-871-64	Sequence 64, Appl
17	34	63.0	676	3	US-09-313-930-2	Sequence 2, Appli
18	33	61.1	119	4	US-09-107-532A-4666	Sequence 4666, Ap
19	33	61.1	158	2	US-08-729-103-1	Sequence 1, Appli
20	33	61.1	158	2	US-08-468-413-2	Sequence 2, Appli
21	33	61.1	158	3	US-09-162-508-2	Sequence 2, Appli
22	33	61.1	158	5	PCT-US95-07169-2	Sequence 2, Appli
23	33	61.1	727	4	US-09-252-991A-18812	Sequence 18812, A
24	32	59.3	50	5	PCT-US93-05640-37	Sequence 37, Appl
25	32	59.3	101	4	US-09-732-210-250	Sequence 250, App
26	32	59.3	175	4	US-09-252-991A-23113	Sequence 23113, A
27	32	59.3	445	4	US-09-252-991A-20277	Sequence 20277, A
28	32	59.3	651	1	US-08-264-101-2	Sequence 2, Appli
29	32	59.3	651	2	US-08-765-243-2	Sequence 2, Appli
30	32	59.3	651	5	PCT-US95-07295-2	Sequence 2, Appli
31	32	59.3	734	2	US-08-765-243-8	Sequence 8, Appli
32	32	59.3	734	5	PCT-US95-07295-8	Sequence 8, Appli
33	31	57.4	154	4	US-09-252-991A-31454	Sequence 31454, A
34	31	57.4	326	2	US-09-055-097-6	Sequence 6, Appli
35	31	57.4	450	4	US-09-252-991A-32284	Sequence 32284, A
36	31	57.4	722	4	US-09-617-145-2	Sequence 2, Appli
37	31	57.4	1196	4	US-09-252-991A-18261	Sequence 18261, A
38	31	57.4	1732	2	US-08-477-451-14	Sequence 14, Appl
39	30.5	56.5	220	4	US-09-252-991A-24410	Sequence 24410, A
40	30	55.6	25	1	US-08-540-202-1	Sequence 1, Appli
41	30	55.6	25	4	US-09-750-726-1	Sequence 1, Appli
42	30	55.6	38	1	US-08-540-202-2	Sequence 2, Appli
43	30	55.6	38	4	US-09-750-726-2	Sequence 2, Appli
44	30	55.6	40	4	US-09-461-325-314	Sequence 314, App
45	30	55.6	50	3	US-09-230-637-56	Sequence 56, Appl

ALIGNMENTS

RESULT 1

US-08-526-710-1

; Sequence 1, Application US/08526710

; Patent No. 5622699

; GENERAL INFORMATION:

; APPLICANT: Ruoslahti, Erkki

; APPLICANT: Pasqualini, Renata

; TITLE OF INVENTION: Method of Identifying Molecules That

; TITLE OF INVENTION: Home to a Selected Organ In Vivo

; NUMBER OF SEQUENCES: 44

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell and Flores

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego

; STATE: California

; COUNTRY: United States

; ZIP: 92122

; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,710
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-526-710-1

```

```

Query Match      100.0%; Score 54; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      9; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 CNSRLHLRC 9
        |||||
Db      1 CNSRLHLRC 9

```

RESULT 2

```

US-08-862-855-1
; Sequence 1, Application US/08862855
; Patent No. 6068829
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/862,855
; FILING DATE:

```

```

; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2621
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-862-855-1

```

```

Query Match          100.0%; Score 54; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 CNSRLHLRC 9
        |||||
Db      1 CNSRLHLRC 9

```

RESULT 3

US-09-226-985-1

```

; Sequence 1, Application US/09226985
; Patent No. 6296832
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Molecules That Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/226,985
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```



```

; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-226-985-1

```

```

Query Match          100.0%; Score 54; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      9; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 CNSRLHLRC 9
        |||||
Db      1 CNSRLHLRC 9

```

RESULT 4

```

US-09-227-906-1
; Sequence 1, Application US/09227906
; Patent No. 6306365
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/227,906
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3424
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-227-906-1

```

```

Query Match          100.0%; Score 54; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      9; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 CNSRLHLRC 9
        |||||
Db      1 CNSRLHLRC 9

```

RESULT 5

US-08-526-710-5

```

; Sequence 5, Application US/08526710
; Patent No. 5622699
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/526,710
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-526-710-5

```

```

Query Match          85.2%; Score 46; DB 1; Length 9;
Best Local Similarity 88.9%; Pred. No. 2.5e+05;
Matches      8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 CNSRLHLRC 9
        ||||| |||
Db      1 CNSRLQLRC 9

```

RESULT 6

US-08-862-855-5

```

; Sequence 5, Application US/08862855
; Patent No. 6068829
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/862,855
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2621
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-862-855-5

```

```

Query Match          85.2%; Score 46; DB 3; Length 9;
Best Local Similarity 88.9%; Pred. No. 2.5e+05;
Matches      8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 CNSRLHLC 9
        ||||| |||
Db      1 CNSRLQLRC 9

```

RESULT 7

US-09-226-985-5

```

; Sequence 5, Application US/09226985
; Patent No. 6296832
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Molecules That Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/226,985
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-226-985-5

```

```

Query Match          85.2%; Score 46; DB 3; Length 9;
Best Local Similarity 88.9%; Pred. No. 2.5e+05;
Matches      8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 CNSRLHLRC 9
        ||||| |||
Db      1 CNSRLQLRC 9

```

```

RESULT 8
US-09-227-906-5
; Sequence 5, Application US/09227906
; Patent No. 6306365
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/227,906
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3424
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-227-906-5

```

```

Query Match          85.2%; Score 46; DB 4; Length 9;
Best Local Similarity 88.9%; Pred. No. 2.5e+05;
Matches      8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 CNSRLHLRC 9
        ||||| |||
Db      1 CNSRLQLRC 9

```

RESULT 9

US-09-738-946-8

```

; Sequence 8, Application US/09738946
; Patent No. 6579701
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: DROSOPHILA HOMOLOGUES OF GENES AND PROTEINS IMPLICATED
IN CANCER AND
; TITLE OF INVENTION: METHODS OF USE
; FILE REFERENCE: EX00-043C
; CURRENT APPLICATION NUMBER: US/09/738,946
; CURRENT FILING DATE: 2000-12-14
; PRIOR APPLICATION NUMBER: 60/170,832
; PRIOR FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: 60/170,838
; PRIOR FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: 60/178,580
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/185,879
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: 60/185,880
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: 60/186,150
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/189,701
; PRIOR FILING DATE: 2000-03-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0

```

; SEQ ID NO 8
; LENGTH: 781
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-738-946-8

Query Match 66.7%; Score 36; DB 4; Length 781;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
| | | | |
Db 529 CNSRGHCHC 537

RESULT 10
US-09-252-991A-28422
; Sequence 28422, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28422
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28422

Query Match 64.8%; Score 35; DB 4; Length 397;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 RLHLRC 9
| | | | |
Db 146 RLHLRC 151

RESULT 11
US-09-252-991A-28223
; Sequence 28223, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28223
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28223

Query Match 64.8%; Score 35; DB 4; Length 511;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
| | | | |
Db 130 CPSRTHRRRC 138

RESULT 12

US-09-330-740A-8

; Sequence 8, Application US/09330740A
; Patent No. 6291217
; GENERAL INFORMATION:
; APPLICANT: Floh , Leopold
; APPLICANT: Koenig, Kerstin
; APPLICANT: Menge, Ulrich
; TITLE OF INVENTION: Glutathionylspermidine Synthetase and
; TITLE OF INVENTION: Processes for Recovery and Use Thereof
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/ 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/330,740A
; FILING DATE: 11-JUN-1999
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/EP97/06982
; FILING DATE: 12-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, James P.
; REGISTRATION NUMBER: 28,491
; REFERENCE/DOCKET NUMBER: 29473/35677
; TELECOMMUNICATION INFORMATION:


```

; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 573 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 191
; OTHER INFORMATION: /note= "Xaa = Lys or Asn"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 463
; OTHER INFORMATION: /note= "Xaa = Val or Asp"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 479
; OTHER INFORMATION: /note= "Xaa = Val or Gly"
US-09-330-740A-8

```

```

Query Match          64.8%; Score 35; DB 3; Length 573;
Best Local Similarity 55.6%; Pred. No. 1.6e+02;
Matches      5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy      1 CNSRLHLRC 9
        |: ||||
Db      236 CDHEPHLRC 244

```

RESULT 13

```

US-09-004-838-11
; Sequence 11, Application US/09004838
; Patent No. 6350933
; GENERAL INFORMATION:
; APPLICANT: Micheltmore, Richard W.
; APPLICANT: Shen, Kathy
; APPLICANT: Meyers, Blake
; TITLE OF INVENTION: Procedures and Materials for
; TITLE OF INVENTION: Conferring Pest Resistance in Plants
; NUMBER OF SEQUENCES: 140
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/004,838
; FILING DATE: 09-JAN-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/781,734
; FILING DATE: 10-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Einhorn, Gregory P.
; REGISTRATION NUMBER: 38,440
; REFERENCE/DOCKET NUMBER: 023070-078810US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1402 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..1402
; OTHER INFORMATION: /note= "RLG1A amino acids"
US-09-004-838-11

```

```

Query Match          64.8%; Score 35; DB 4; Length 1402;
Best Local Similarity 66.7%; Pred. No. 3.8e+02;
Matches      6; Conservative      0; Mismatches      3; Indels      0; Gaps      0;

```

```

Qy      1 CNSRLHLRC 9
        ||| | ||
Db      1264 CNSLEHCRC 1272

```

```

RESULT 14
US-09-082-358B-88
; Sequence 88, Application US/09082358B
; Patent No. 6469153
; GENERAL INFORMATION:
; APPLICANT: Goff, Stephen P.
; APPLICANT: Li, Xingquiang
; TITLE OF INVENTION: EIP-1, EIP-3 GENES, ENVELOPE-INTERACTING PROTEINS,
; TITLE OF INVENTION: EIP-1, and EIP-3
; FILE REFERENCE: 0575/54804
; CURRENT APPLICATION NUMBER: US/09/082,358B
; CURRENT FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 88
; LENGTH: 36
; TYPE: PRT
; ORGANISM: murine
US-09-082-358B-88

```

```

Query Match          63.0%; Score 34; DB 4; Length 36;

```

Best Local Similarity 85.7%; Pred. No. 15;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 SRLHLRC 9
|| ||||
Db 3 SRTHLRC 9

RESULT 15

US-09-227-357-580

; Sequence 580, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: PZ010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948

```

; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,950
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,947
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,964
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 580
;   LENGTH: 40
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-227-357-580

```

```

Query Match          63.0%;   Score 34;   DB 4;   Length 40;
Best Local Similarity 44.4%;   Pred. No. 17;
Matches      4;   Conservative      4;   Mismatches      1;   Indels      0;   Gaps      0;

```

```

Qy      1 CNSRLHLRC 9
        | :|:|::|
Db      2 CVTRMHVKC 10

```

Search completed: November 13, 2003, 09:54:55
Job time : 11.6875 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:46:46 ; Search time 152.438 Seconds
(without alignments)
53.722 Million cell updates/sec

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 30.2812 Seconds
(without alignments)
47.176 Million cell updates/sec

Title: US-09-228-866-1
Perfect score: 54
Sequence: 1 CNSRLHLRC 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*

- 1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
- 2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
- 3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
- 4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
- 5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
- 6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
- 7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
- 8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
- 9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
- 10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
- 11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
- 12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
- 13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
- 14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
- 15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
- 16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
- 17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
- 18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
- 19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
- 20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
- 21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
- 22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
- 23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
- 24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result		%					
No.	Score	Query	Match	Length	DB	ID	Description
1	54	100.0		9	18	AAW13410	Brain homing pepti
2	54	100.0		9	21	AAB07387	Brain homing pepti
3	54	100.0		9	22	AAE11793	Phage peptide #1 t
4	54	100.0		9	23	AAU10704	Brain homing pepti
5	54	100.0		9	24	ABU59529	Brain receptor tar
6	46	85.2		9	18	AAW13411	Brain homing pepti
7	46	85.2		9	21	AAB07391	Brain homing pepti
8	46	85.2		9	22	AAE11797	Phage peptide #5 t
9	46	85.2		9	23	AAU10708	Brain homing pepti
10	40	74.1		341	22	ABG22110	Novel human diagno
11	40	74.1		1327	21	AAV70474	Human cyclic nucle
12	39	72.2		1091	22	ABB58383	Drosophila melanog
13	38	70.4		167	21	AAG16491	Arabidopsis thalia
14	38	70.4		167	21	AAG37995	Arabidopsis thalia
15	38	70.4		199	21	AAG19731	Arabidopsis thalia
16	38	70.4		220	21	AAG16490	Arabidopsis thalia
17	38	70.4		220	21	AAG37994	Arabidopsis thalia
18	38	70.4		247	21	AAG16489	Arabidopsis thalia
19	38	70.4		247	21	AAG37993	Arabidopsis thalia
20	38	70.4		251	21	AAG19730	Arabidopsis thalia
21	38	70.4		315	23	ABB84457	Siah-related prote
22	38	70.4		739	22	ABB59471	Drosophila melanog
23	37	68.5		36	22	ABG07364	Novel human diagno
24	36	66.7		84	20	AAV60189	Human endometrium
25	36	66.7		93	22	ABG26573	Novel human diagno
26	36	66.7		129	22	ABG12071	Novel human diagno
27	36	66.7		191	22	AAU87130	Novel central nerv
28	36	66.7		414	23	ABB89967	Human polypeptide
29	36	66.7		449	22	AAU19452	Human diagnostic a
30	36	66.7		468	22	ABB65114	Drosophila melanog
31	36	66.7		522	22	AAU19470	Human diagnostic a
32	36	66.7		1234	23	ABP68968	Human polypeptide
33	36	66.7		1569	22	ABB63422	Drosophila melanog
34	35	64.8		51	23	ABP32783	Human ORF1756 prot
35	35	64.8		104	22	AAB95889	Human protein sequ
36	35	64.8		140	21	AAG36307	Arabidopsis thalia
37	35	64.8		144	21	AAG36306	Arabidopsis thalia
38	35	64.8		175	21	AAG36305	Arabidopsis thalia
39	35	64.8		261	21	AAG36578	Arabidopsis thalia
40	35	64.8		289	22	ABG26084	Novel human diagno
41	35	64.8		305	18	AAW55451	H. pylori ORF 02ae
42	35	64.8		306	19	AAW98321	H. pylori GHPO 134
43	35	64.8		306	19	AAW71512	Helicobacter polyp
44	35	64.8		321	21	AAG13847	Arabidopsis thalia
45	35	64.8		321	21	AAG53803	Arabidopsis thalia

ALIGNMENTS

```

RESULT 1
AAW13410
ID  AAW13410 standard; Peptide; 9 AA.
XX
AC  AAW13410;
XX
DT  15-JAN-1998  (first entry)
XX
DE  Brain homing peptide.
XX
KW  Brain homing peptide; in vivo panning; screening; phage display.
XX
OS  Synthetic.
XX
PN  WO9710507-A1.
XX
PD  20-MAR-1997.
XX
PF  10-SEP-1996;   96WO-US14600.
XX
PR  11-SEP-1995;   95US-0526710.
PR  11-SEP-1995;   95US-0526708.
XX
PA  (LJOL-) LA JOLLA CANCER RES FOUND.
XX
PI  Pasqualini R,  Ruoslahti E;
XX
DR  WPI; 1997-202359/18.
XX
PT  Obtaining compound that homes to selected organ or tissue - by in
PT  vivo panning method, specifically to identify brain, kidney,
PT  angiogenic vasculature or tumour tissue homing peptide(s)
XX
PS  Claim 11; Page 67; 75pp; English.
XX
CC  This synthetic peptide is a claimed example of a brain-homing
CC  peptide that was identified using a novel method for obtaining
CC  molecules that home to a selected organ or tissue. This in vivo
CC  panning method typically involves administering a phage display
CC  library to a subject, and identifying expressed peptides which
CC  home to the desired organ or tissue, e.g. brain, kidney, angiogenic
CC  vascular tissue or tumour tissue. The isolated peptides (see
CC  AAW13412-52, AAW11181-86) can be used to target e.g. drugs, toxins or
CC  labels to the selected organ/tissue (claimed) or to identify and/or
CC  isolate target molecules (claimed). The peptides can be directly
CC  identified in vivo, as compared to prior art in vitro screening
CC  methods, which require further examination to see if they maintain
CC  specificity in vivo.
XX
SQ  Sequence    9 AA;

Query Match          100.0%;  Score 54;  DB 18;  Length 9;
Best Local Similarity 100.0%;  Pred. No. 9.3e+05;
Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

```

```

Qy          1 CNSRLHLRC 9
            |||||

```

Db 1 CNSRLHLRC 9

RESULT 2

AAB07387

ID AAB07387 standard; peptide; 9 AA.

XX

AC AAB07387;

XX

DT 17-OCT-2000 (first entry)

XX

DE Brain homing peptide # 1.

XX

KW Brain; homing peptide; organ targeting; tissue targeting; mouse; cyclic.

XX

OS Mus sp.

XX

FH Key Location/Qualifiers

FT Disulfide-bond 1..9

FT /note= "Can optionally form a cyclic peptide"

XX

PN US6068829-A.

XX

PD 30-MAY-2000.

XX

PF 23-JUN-1997; 97US-0862855.

XX

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Pasqualini R, Ruoslahti E;

XX

DR WPI; 2000-410850/35.

XX

PT Identifying and recovering organ homing molecules or peptides by in
PT vivo panning comprises administering a library of diverse peptides
PT linked to a tag which facilitates recovery of these peptides -

XX

PS Example 2; Column 17; 20pp; English.

XX

CC The present sequence is a mouse brain homing peptide. This sequence was
CC identified by using in vivo panning to screen a library of potential
CC organ homing molecules. The present sequence can be used to direct a
CC moiety to a the brain tissue, by linking the moiety to the present
CC sequence. Examples of potential moieties are drugs, toxins or a
CC detectable label. The present sequence contains a SRL amino acid motif.

XX

SQ Sequence 9 AA;

Query Match 100.0%; Score 54; DB 21; Length 9;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9

|||||||

Db

1 CNSRLHLRC 9

RESULT 3

AAE11793

ID AAE11793 standard; peptide; 9 AA.

XX

AC AAE11793;

XX

DT 18-DEC-2001 (first entry)

XX

DE Phage peptide #1 targetted to brain.

XX

KW Enriched library fraction; brain; kidney; tumour; panning; diagnostic;
KW molecular medicine; drug delivery; peptidomimetic; pharmaceutical.

XX

OS Bacteriophage.

XX

FH Key Location/Qualifiers

FT Domain 3..5

FT /label= SRL_motif

XX

PN US6296832-B1.

XX

PD 02-OCT-2001.

XX

PF 08-JAN-1999; 99US-0226985.

XX

PR 23-JUN-1997; 97US-0862855.

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ruoslahti E, Pasqualini R;

XX

DR WPI; 2001-610691/70.

XX

PT Enriched library fraction comprising molecules recovered by in vivo
PT panning that selectively home to a selected organ or tissue useful for
PT treating disease or in diagnostic methods -

XX

PS Example 2; Column 17; 21pp; English.

XX

CC The invention relates to an enriched library fraction containing
CC molecules that selectively home to a selected organ or tissue such as
CC brain, kidney or tumour recovered by in vivo panning. The invention
CC generally relates to the field of molecular medicine, drug delivery and
CC to a method of invivo panning for identifying a molecule that homes to a
CC specific organ. The molecules, e.g., peptides, peptidomimetics, proteins
CC and fragments of proteins contained in an enriched library fraction may
CC be administered to a subject as part of a pharmaceutical composition to
CC treat disease or in diagnostic methods. The present sequence is a
CC peptide from bacteriophage targetted to brain.

XX

SQ Sequence 9 AA;

Query Match 100.0%; Score 54; DB 22; Length 9;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
|||
Db 1 CNSRLHLRC 9

RESULT 4

AAU10704

ID AAU10704 standard; peptide; 9 AA.

XX

AC AAU10704;

XX

DT 12-MAR-2002 (first entry)

XX

DE Brain homing peptide #1 useful for delivery of target molecules.

XX

KW Organ targeting; tissue targeting; cancer; tumour homing molecule;
KW delivery of target molecule; brain homing peptide.

XX

OS Synthetic.

XX

PN US6306365-B1.

XX

PD 23-OCT-2001.

XX

PF 08-JAN-1999; 99US-0227906.

XX

PR 23-JUN-1997; 97US-0862855.

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ruoslahti E, Pasqualini R;

XX

DR WPI; 2002-040196/05.

XX

PT Recovering molecules that home to an organ or tissue, useful for
PT identifying molecules that home to a specific organ or tissue, e.g.
PT identifying a tumour homing molecule to identify the presence of cancer,
PT by in vivo panning of a library -

XX

PS Example 2; Column 17; 21pp; English.

XX

CC The present invention relates to a method of recovering molecules that
CC home to a selected organ or tissue. The method comprises administering
CC to the subject the library of diverse molecules, collecting a sample of
CC the selected organ or tissue (e.g. brain or kidney), and recovering from
CC the sample several molecules that home to the selected organ or tissue.
CC The method is useful for identifying molecules, particularly useful for
CC screening large number of molecules (e.g. peptides), that home to a
CC specific organ. The identified molecule is useful for e.g. raising an
CC antibody specific for a target molecule, targeting a desired moiety
CC (e.g. drug, toxin or detectable label) to the selected organ.

CC Specifically, the method is useful for identifying the presence of cancer
CC in a subject by linking an appropriate moiety to a tumour homing
CC molecule. The present method provides a direct means for identifying
CC molecules that specifically home to a selected organ and, therefore
CC provides a significant advantage over previous methods, which require
CC that a molecule identified using an in vitro screening method
CC subsequently be examined to determine if it maintains its specificity in
CC vivo. AAU10704-AAU10723 represent brain homing peptides described in
CC the present invention.

XX

SQ Sequence 9 AA;

Query Match 100.0%; Score 54; DB 23; Length 9;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9

|||||||

Db 1 CNSRLHLRC 9

RESULT 5

ABU59529

ID ABU59529 standard; Peptide; 9 AA.

XX

AC ABU59529;

XX

DT 22-APR-2003 (first entry)

XX

DE Brain receptor targeting peptide #1.

XX

KW Targeting ligand; bioactive agent; polymer matrix; cancer; cytostatic;

KW cathepsin-D substrate; peptides; neuroreceptor; adrenal receptor;

KW fibronectin; vitronectin; integrin; RGD motif; angiogenic endothelium;

KW tumour; cationic cancer-targeting peptide.

XX

OS Synthetic.

XX

PN US2002041898-A1.

XX

PD 11-APR-2002.

XX

PF 25-JUL-2001; 2001US-0912609.

XX

PR 05-JAN-2000; 2000US-0478124.

PR 31-OCT-2000; 2000US-0703474.

XX

PA (UNGE/) UNGER E C.

PA (MATS/) MATSUNAGA T O.

PA (RAMA/) RAMASWAMI V.

PA (ROMA/) ROMANOWSKI M J.

XX

PI Unger EC, Matsunaga TO, Ramaswami V, Romanowski MJ;

XX

DR WPI; 2003-208921/20.

XX

PT Targeted delivery system comprising a bioactive agent homogeneously

PT dispersed in a targeted matrix is especially useful in cancer therapy
PT -

XX

PS Claim 23; Page 37; 46pp; English.

XX

CC The invention relates to a composition comprising a bioactive agent
CC homogeneously dispersed in a targeted matrix (polymer and targeting
CC ligand). Also included are a targeted matrix for use as a delivery
CC vehicle comprising a polymer associated with a targeting ligand,
CC enhancing the bioavailability of an agent comprising administration
CC of the composition and treating cancer comprising administration of the
CC novel composition. The method is useful for targeted delivery of a drug,
CC especially in cancer therapy. The targeting ligand may be a peptide.
CC Examples of targeting peptides are disclosed including cathepsin-D
CC substrate peptides, peptides targeting receptors in the brain and
CC kidney, peptides recognising fibronectin- and vitronectin-binding
CC integrins, peptides targeting the RGD (Arg-Gly-Asp)-motif in, e.g.,
CC antibodies, peptides targeting the angiogenic endothelium of solid
CC tumours, tissue specific peptides (e.g. of lung, skin, pancreas,
CC intestine, uterus, adrenal gland and retina), and cationic cancer-
CC targeting peptides. The present sequence is a peptide targeting
CC ligand disclosed in the invention.

XX

SQ Sequence 9 AA;

Query Match 100.0%; Score 54; DB 24; Length 9;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
|||
Db 1 CNSRLHLRC 9

RESULT 6

AAW13411

ID AAW13411 standard; Peptide; 9 AA.

XX

AC AAW13411;

XX

DT 15-JAN-1998 (first entry)

XX

DE Brain homing peptide.

XX

KW Brain homing peptide; in vivo panning; screening; phage display;
KW drug delivery.

XX

OS Synthetic.

XX

PN WO9710507-A1.

XX

PD 20-MAR-1997.

XX

PF 10-SEP-1996; 96WO-US14600.

XX

PR 11-SEP-1995; 95US-0526710.

PR 11-SEP-1995; 95US-0526708.

XX
PA (LJOL-) LA JOLLA CANCER RES FOUND.
XX
PI Pasqualini R, Ruoslahti E;
XX
DR WPI; 1997-202359/18.
XX
PT Obtaining compound that homes to selected organ or tissue - by in
PT vivo panning method, specifically to identify brain, kidney,
PT angiogenic vasculature or tumour tissue homing peptide(s)
XX
PS Claim 11; Page 67; 75pp; English.
XX
CC This synthetic peptide is a claimed example of a brain-homing
CC peptide that was identified using a novel method for obtaining
CC molecules that home to a selected organ or tissue. This in vivo
CC panning method typically involves administering a phage display
CC library to a subject, and identifying expressed peptides which
CC home to the desired organ or tissue, e.g. brain, kidney, angiogenic
CC vascular tissue or tumour tissue. The isolated peptides (see
CC AAW13412-52, AAW11181-86) can be used to target e.g. drugs, toxins or
CC labels to the selected organ/tissue (claimed) or to identify and/or
CC isolate target molecules (claimed). The peptides can be directly
CC identified in vivo, as compared to prior art in vitro screening
CC methods, which require further examination to see if they maintain
CC specificity in vivo.
XX
SQ Sequence 9 AA;

Query Match 85.2%; Score 46; DB 18; Length 9;
Best Local Similarity 88.9%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
||||| |||
Db 1 CNSRLQLRC 9

RESULT 7

AAB07391
ID AAB07391 standard; peptide; 9 AA.
XX
AC AAB07391;
XX
DT 17-OCT-2000 (first entry)
XX
DE Brain homing peptide # 5.
XX
KW Brain; homing peptide; organ targeting; tissue targeting; mouse; cyclic.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT Disulfide-bond 1..9
FT /note= "Can optionally form a cyclic peptide"
XX
PN US6068829-A.

XX
 PD 30-MAY-2000.
 XX
 PF 23-JUN-1997; 97US-0862855.
 XX
 PR 11-SEP-1995; 95US-0526710.
 PR 10-MAR-1997; 97US-0813273.
 XX
 PA (BURN-) BURNHAM INST.
 XX
 PI Pasqualini R, Ruoslahti E;
 XX
 DR WPI; 2000-410850/35.
 XX
 PT Identifying and recovering organ homing molecules or peptides by in
 PT vivo panning comprises administering a library of diverse peptides
 PT linked to a tag which facilitates recovery of these peptides -
 XX
 PS Example 2; Column 17; 20pp; English.
 XX
 CC The present sequence is a mouse brain homing peptide. This sequence was
 CC identified by using in vivo panning to screen a library of potential
 CC organ homing molecules. The present sequence can be used to direct a
 CC moiety to a the brain tissue, by linking the moiety to the present
 CC sequence. Examples of potential moieties are drugs, toxins or a
 CC detectable label. The present sequence contains a SRL amino acid motif.
 XX
 SQ Sequence 9 AA;

Query Match 85.2%; Score 46; DB 21; Length 9;
 Best Local Similarity 88.9%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
 ||||| |||
 Db 1 CNSRLQLRC 9

RESULT 8

AAE11797

ID AAE11797 standard; peptide; 9 AA.

XX

AC AAE11797;

XX

DT 18-DEC-2001 (first entry)

XX

DE Phage peptide #5 targetted to brain.

XX

KW Enriched library fraction; brain; kidney; tumour; panning; diagnostic;
 KW molecular medicine; drug delivery; peptidomimetic; pharmaceutical.

XX

OS Bacteriophage.

XX

FH Key Location/Qualifiers

FT Domain 3..5

FT /label= SRL_motif

XX

PN US6296832-B1.
 XX
 PD 02-OCT-2001.
 XX
 PF 08-JAN-1999; 99US-0226985.
 XX
 PR 23-JUN-1997; 97US-0862855.
 PR 11-SEP-1995; 95US-0526710.
 PR 10-MAR-1997; 97US-0813273.
 XX
 PA (BURN-) BURNHAM INST.
 XX
 PI Ruoslahti E, Pasqualini R;
 XX
 DR WPI; 2001-610691/70.
 XX
 PT Enriched library fraction comprising molecules recovered by in vivo
 PT panning that selectively home to a selected organ or tissue useful for
 PT treating disease or in diagnostic methods -
 XX
 PS Example 2; Column 17; 21pp; English.
 XX
 CC The invention relates to an enriched library fraction containing
 CC molecules that selectively home to a selected organ or tissue such as
 CC brain, kidney or tumour recovered by in vivo panning. The invention
 CC generally relates to the field of molecular medicine, drug delivery and
 CC to a method of invivo panning for identifying a molecule that homes to a
 CC specific organ. The molecules, e.g., peptides, peptidomimetics, proteins
 CC and fragments of proteins contained in an enriched library fraction may
 CC be administered to a subject as part of a pharmaceutical composition to
 CC treat disease or in diagnostic methods. The present sequence is a
 CC peptide from bacteriophage targetted to brain.
 XX
 SQ Sequence 9 AA;

 Query Match 85.2%; Score 46; DB 22; Length 9;
 Best Local Similarity 88.9%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
 ||||| |||
 Db 1 CNSRLQLRC 9

RESULT 9
 AAU10708
 ID AAU10708 standard; peptide; 9 AA.
 XX
 AC AAU10708;
 XX
 DT 12-MAR-2002 (first entry)
 XX
 DE Brain homing peptide #5 useful for delivery of target molecules.
 XX
 KW Organ targeting; tissue targeting; cancer; tumour homing molecule;
 KW delivery of target molecule; brain homing peptide.
 XX

OS Synthetic.
XX
PN US6306365-B1.
XX
PD 23-OCT-2001.
XX
PF 08-JAN-1999; 99US-0227906.
XX
PR 23-JUN-1997; 97US-0862855.
PR 11-SEP-1995; 95US-0526710.
PR 10-MAR-1997; 97US-0813273.
XX
PA (BURN-) BURNHAM INST.
XX
PI Ruoslahti E, Pasqualini R;
XX
DR WPI; 2002-040196/05.
XX
PT Recovering molecules that home to an organ or tissue, useful for
PT identifying molecules that home to a specific organ or tissue, e.g.
PT identifying a tumour homing molecule to identify the presence of cancer,
PT by in vivo panning of a library -
XX
PS Example 2; Column 17; 21pp; English.
XX
CC The present invention relates to a method of recovering molecules that
CC home to a selected organ or tissue. The method comprises administering
CC to the subject the library of diverse molecules, collecting a sample of
CC the selected organ or tissue (e.g. brain or kidney), and recovering from
CC the sample several molecules that home to the selected organ or tissue.
CC The method is useful for identifying molecules, particularly useful for
CC screening large number of molecules (e.g. peptides), that home to a
CC specific organ. The identified molecule is useful for e.g. raising an
CC antibody specific for a target molecule, targeting a desired moiety
CC (e.g. drug, toxin or detectable label) to the selected organ.
CC Specifically, the method is useful for identifying the presence of cancer
CC in a subject by linking an appropriate moiety to a tumour homing
CC molecule. The present method provides a direct means for identifying
CC molecules that specifically home to a selected organ and, therefore
CC provides a significant advantage over previous methods, which require
CC that a molecule identified using an in vitro screening method
CC subsequently be examined to determine if it maintains its specificity in
CC vivo. AAU10704-AAU10723 represent brain homing peptides described in
CC the present invention.
XX
SQ Sequence 9 AA;

Query Match 85.2%; Score 46; DB 23; Length 9;
Best Local Similarity 88.9%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
||||| |||
Db 1 CNSRLQLRC 9

ABG22110

ID ABG22110 standard; Protein; 341 AA.

XX

AC ABG22110;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #22101.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

PR

23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

DR

N-PSDB; AAS86297.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX

PS Claim 20; SEQ ID No 52469; 103pp; English.

XX

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 341 AA;

Query Match 74.1%; Score 40; DB 22; Length 341;

Best Local Similarity 66.7%; Pred. No. 65;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9

|: |||||

Db 150 CSGHLHLRC 158

RESULT 11

AAAY70474

ID AAY70474 standard; Protein; 1327 AA.

XX

AC AAY70474;

XX

DT 04-JUL-2000 (first entry)

XX

DE Human cyclic nucleotide-associated protein-2 (CNAP-2).

XX

KW Cyclic nucleotide-associated protein-2; CNAP-2; human; cytostatic;
KW anti-arteriosclerotic; hepatotropic; anti-leukaemic; anti-inflammatory;
KW immunomodulatory; anti-asthmatic; anti-anaemic; anti-diabetic; diagnosis;
KW anti-sclerotic; dermatological; neuroprotective; anti-epileptic; cancer;
KW anti-Alzheimer's; anti-Parkinsonian; cerebroprotective; ophthalmological;
KW anti-infertility; anti-allergic; vasotropic; immunosuppressive;
KW hypotensive; gene therapy; prevention; treatment; arteriosclerosis;
KW cell proliferative disorder; autoimmune/inflammatory; diabetes mellitus;
KW neurological; vision; reproductive; smooth muscle.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Peptide 1..34

FT /label= Signal_peptide

FT Protein 35..1327

FT /label= Mature_CNAP-2

FT /note= "Shares 24% identity to Aquifex pyrophilus

FT esterase 28LC"

FT Modified-site 68

FT /note= "Potential phosphorylation site"

FT Modified-site 1225

FT /note= "Potential phosphorylation site"

FT Modified-site 73

FT /note= "Potential phosphorylation site"

FT Modified-site 125

FT /note= "Potential phosphorylation site"

FT Modified-site 220

FT /note= "Potential phosphorylation site"

FT Modified-site 326

FT /note= "Potential phosphorylation site"

FT Modified-site 357

FT /note= "Potential phosphorylation site"

FT Modified-site 386

FT		/note= "Potential phosphorylation site"
FT	Modified-site	400
FT		/note= "Potential phosphorylation site"
FT	Modified-site	432
FT		/note= "Potential phosphorylation site"
FT	Modified-site	455
FT		/note= "Potential phosphorylation site"
FT	Modified-site	560
FT		/note= "Potential phosphorylation site"
FT	Modified-site	600
FT		/note= "Potential phosphorylation site"
FT	Modified-site	780
FT		/note= "Potential phosphorylation site"
FT	Modified-site	784
FT		/note= "Potential phosphorylation site"
FT	Modified-site	997
FT		/note= "Potential phosphorylation site"
FT	Modified-site	1113
FT		/note= "Potential phosphorylation site"
FT	Modified-site	1121
FT		/note= "Potential phosphorylation site"
FT	Modified-site	1171
FT		/note= "Potential phosphorylation site"
FT	Modified-site	1251
FT		/note= "Potential phosphorylation site"
FT	Modified-site	1274
FT		/note= "Potential phosphorylation site"
FT	Modified-site	1285
FT		/note= "Potential phosphorylation site"
FT	Modified-site	1299
FT		/note= "Potential phosphorylation site"
FT	Modified-site	1301
FT		/note= "Potential phosphorylation site"
FT	Modified-site	1323
FT		/note= "Potential phosphorylation site"
FT	Modified-site	82
FT		/note= "Potential phosphorylation site"
FT	Modified-site	236
FT		/note= "Potential phosphorylation site"
FT	Modified-site	319
FT		/note= "Potential phosphorylation site"
FT	Modified-site	547
FT		/note= "Potential phosphorylation site"
FT	Modified-site	634
FT		/note= "Potential phosphorylation site"
FT	Modified-site	699
FT		/note= "Potential phosphorylation site"
FT	Modified-site	816
FT		/note= "Potential phosphorylation site"
FT	Modified-site	894
FT		/note= "Potential phosphorylation site"
FT	Modified-site	910
FT		/note= "Potential phosphorylation site"
FT	Modified-site	1220
FT		/note= "Potential phosphorylation site"
FT	Modified-site	1230
FT		/note= "Potential phosphorylation site"

FT Modified-site 392
 FT /note= "Potential phosphorylation site"
 FT Modified-site 1019
 FT /note= "N-glycosylated"
 FT Modified-site 1040
 FT /note= "N-glycosylated"
 FT Modified-site 1228
 FT /note= "N-glycosylated"
 FT Binding-site 144..269
 FT /label= cNMP-binding_domain
 FT Binding-site 573..696
 FT /label= cNMP-binding_domain
 FT Domain 10..30
 FT /label= Transmembrane_domain
 FT Region 605..628
 FT /note= "Resembles cyclic-nucleotide binding domain
 FT proteins"
 FT Region 643..676
 FT /note= "Resembles cyclic-nucleotide binding domain
 FT proteins"
 XX
 PN WO200014248-A1.
 XX
 PD 16-MAR-2000.
 XX
 PF 03-SEP-1999; 99WO-US20287.
 XX
 PR 04-SEP-1998; 98US-0148904.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Hillman JL, Yue H, Guegler KJ, Corley NC, Patterson C, Tang YT;
 XX
 DR WPI; 2000-256994/22.
 DR N-PSDB; AAZ51683.
 XX
 PT Isolated cyclic nucleotide associated proteins useful for preventing,
 PT diagnosing and treating cell proliferative, autoimmune/inflammatory,
 PT neurological, vision, reproductive and smooth muscle disorders -
 XX
 PS Disclosure; Page 64-67; 78pp; English.
 XX
 CC The present sequence is a human cyclic nucleotide-
 CC associated protein-2 (CNAP-2), identified in Incyte clone 3149674,
 CC that is isolated from ADRENON04 cDNA library. It is expressed in
 CC nervous, reproductive, cardiovascular and haematopoietic/immune tissues.
 CC CNAP sequences may be used for prevention, treatment and diagnosis of
 CC diseases associated with altered CNAP expression such as, cell
 CC proliferative disorders (e.g. arteriosclerosis, cirrhosis, leukaemia,
 CC lymphoma and cancer of the breast, prostate, lung and brain), autoimmune/
 CC inflammatory disorders (e.g. asthma, anaemia, diabetes mellitus, multiple
 CC sclerosis and psoriasis), neurological disorders (e.g. epilepsy,
 CC Alzheimer's/Parkinson's disease and strokes), vision disorders (e.g.
 CC conjunctivitis, glaucoma, cataracts and retinitis pigmentosa),
 CC reproductive disorders (e.g. infertility, uterine fibroids, ectopic
 CC pregnancies and impotence) and smooth muscle disorders (e.g. angina,
 CC anaphylactic shock, Kearns-Sayre syndrome and hypertension). The

CC coding sequence can be used for gene therapy.

XX

SQ Sequence 1327 AA;

Query Match 74.1%; Score 40; DB 21; Length 1327;

Best Local Similarity 66.7%; Pred. No. 2.2e+02;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9

|: |||||

Db 880 CSGHLHLRC 888

RESULT 12

ABB58383

ID ABB58383 standard; Protein; 1091 AA.

XX

AC ABB58383;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 1941.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX

PA (PEKE) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

DR N-PSDB; ABL02486.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

XX

PS Disclosure; SEQ ID NO 1941; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 1091 AA;

Query Match 72.2%; Score 39; DB 22; Length 1091;

Best Local Similarity 87.5%; Pred. No. 2.7e+02;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 NSRLHLRC 9

Db 307 NFRLHLRC 314

RESULT 13

AAG16491

ID AAG16491 standard; Protein; 167 AA.

XX

AC AAG16491;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 17157.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.

PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.

PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 70.4%; Score 38; DB 21; Length 167;
 Best Local Similarity 66.7%; Pred. No. 74;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CNSRLHLRC 9
 || |||
 Db 96 CNHGFHLRC 104

RESULT 14

AAG37995

ID AAG37995 standard; Protein; 167 AA.

XX

AC AAG37995;

XX

DT 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 46808.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.

PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.

PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.

PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 70.4%; Score 38; DB 21; Length 167;
 Best Local Similarity 66.7%; Pred. No. 74;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
 || ||||
 Db 96 CNHGFHLRC 104

RESULT 15

AAG19731

ID AAG19731 standard; Protein; 199 AA.

XX

AC AAG19731;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 21642.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.

PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.

PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.

PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 70.4%; Score 38; DB 21; Length 199;
 Best Local Similarity 66.7%; Pred. No. 86;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
 | ||:| ||
 Db 18 CKSRVHNRC 26

Search completed: November 13, 2003, 09:45:21
 Job time : 31.2812 secs

GenCore version 5.1.6
 Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:45:35 ; Search time 18.6562 Seconds
 (without alignments)
 88.069 Million cell updates/sec

Title: US-09-228-866-1
 Perfect score: 54
 Sequence: 1 CNSRLHLRC 9

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	54	100.0	9	12	US-10-306-878-11	Sequence 11, Appl
2	35	64.8	306	10	US-09-881-752A-242	Sequence 242, App
3	35	64.8	306	11	US-09-988-067B-78	Sequence 78, Appl
4	35	64.8	378	9	US-09-815-242-10452	Sequence 10452, A
5	35	64.8	387	9	US-09-815-242-5133	Sequence 5133, Ap
6	35	64.8	573	10	US-09-954-433-8	Sequence 8, Appli
7	34.5	63.9	1151	10	US-09-749-601A-10	Sequence 10, Appl
8	34.5	63.9	1151	11	US-09-912-697-33	Sequence 33, Appl
9	34	63.0	40	11	US-09-983-802-580	Sequence 580, App
10	34	63.0	65	12	US-10-029-386-28271	Sequence 28271, A
11	34	63.0	319	10	US-09-771-161A-118	Sequence 118, App
12	34	63.0	676	10	US-09-771-161A-209	Sequence 209, App
13	34	63.0	706	9	US-09-749-956-2	Sequence 2, Appli
14	34	63.0	1154	12	US-10-100-818-4	Sequence 4, Appli
15	33	61.1	29	15	US-10-100-608B-9	Sequence 9, Appli
16	33	61.1	40	11	US-09-852-455-53	Sequence 53, Appl
17	33	61.1	158	9	US-09-922-217-1070	Sequence 1070, Ap
18	33	61.1	158	9	US-09-922-217-1077	Sequence 1077, Ap
19	33	61.1	158	9	US-09-922-217-1078	Sequence 1078, Ap
20	33	61.1	158	9	US-09-922-217-1079	Sequence 1079, Ap
21	33	61.1	158	9	US-09-922-217-1080	Sequence 1080, Ap
22	33	61.1	158	10	US-09-833-263-1070	Sequence 1070, Ap
23	33	61.1	158	10	US-09-833-263-1077	Sequence 1077, Ap
24	33	61.1	158	10	US-09-833-263-1078	Sequence 1078, Ap
25	33	61.1	158	10	US-09-833-263-1079	Sequence 1079, Ap
26	33	61.1	158	10	US-09-833-263-1080	Sequence 1080, Ap
27	33	61.1	158	12	US-09-525-041-2	Sequence 2, Appli
28	33	61.1	158	14	US-10-025-380-1070	Sequence 1070, Ap
29	33	61.1	158	14	US-10-025-380-1077	Sequence 1077, Ap

30	33	61.1	158	14	US-10-025-380-1078	Sequence 1078, Ap
31	33	61.1	158	14	US-10-025-380-1079	Sequence 1079, Ap
32	33	61.1	158	14	US-10-025-380-1080	Sequence 1080, Ap
33	33	61.1	158	15	US-10-100-608B-2	Sequence 2, Appli
34	33	61.1	158	15	US-10-157-031-114	Sequence 114, App
35	33	61.1	158	15	US-10-205-823-347	Sequence 347, App
36	33	61.1	166	15	US-10-106-698-6394	Sequence 6394, Ap
37	33	61.1	169	9	US-09-925-297-746	Sequence 746, App
38	33	61.1	178	15	US-10-106-698-4642	Sequence 4642, Ap
39	33	61.1	227	12	US-10-259-165-100	Sequence 100, App
40	33	61.1	227	12	US-10-259-165-432	Sequence 432, App
41	33	61.1	266	15	US-10-036-542-66	Sequence 66, Appl
42	33	61.1	277	15	US-10-036-542-92	Sequence 92, Appl
43	33	61.1	277	15	US-10-036-542-134	Sequence 134, App
44	33	61.1	288	15	US-10-036-542-130	Sequence 130, App
45	33	61.1	304	10	US-09-925-300-1615	Sequence 1615, Ap

ALIGNMENTS

RESULT 1

US-10-306-878-11

```
; Sequence 11, Application US/10306878
; Publication No. US20030175819A1
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Guo, Bin
; TITLE OF INVENTION: Methods for Identifying Modulators of
; TITLE OF INVENTION: Apoptosis
; FILE REFERENCE: P-LJ 5535
; CURRENT APPLICATION NUMBER: US/10/306,878
; CURRENT FILING DATE: 2002-11-27
; PRIOR APPLICATION NUMBER: US 60/334,149
; PRIOR FILING DATE: 2001-11-28
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
US-10-306-878-11
```

```
Query Match          100.0%; Score 54; DB 12; Length 9;
Best Local Similarity 100.0%; Pred. No. 6e+05;
Matches      9; Conservative    0; Mismatches    0; Indels    0; Gaps    0;
```

```
Qy      1 CNSRLHLRC 9
        |||||
Db      1 CNSRLHLRC 9
```

RESULT 2

US-09-881-752A-242

```
; Sequence 242, Application US/09881752A
```

```

; Patent No. US20020115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides
in the Helicobacter
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/041002
; CURRENT APPLICATION NUMBER: US/09/881,752A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 242
; LENGTH: 306
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-881-752A-242

```

```

Query Match          64.8%; Score 35; DB 10; Length 306;
Best Local Similarity 66.7%; Pred. No. 2.6e+02;
Matches      6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy      1 CNSRLHLRC 9
        || || || |
Db      91 CNLRNHLAC 99

```

RESULT 3

```

US-09-988-067B-78
; Sequence 78, Application US/09988067B
; Publication No. US20030124141A1
; GENERAL INFORMATION:
; APPLICANT: Haas, Rainer
; APPLICANT: Kleanthous, Harold
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Miller, Charles
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Odenbreit, Stefan
; APPLICANT: Meyer, Thomas
; TITLE OF INVENTION: Helicobacter Polypeptides and
; TITLE OF INVENTION: Corresponding Polynucleotide Molecules
; FILE REFERENCE: 06132/040002
; CURRENT APPLICATION NUMBER: US/09/988,067B
; CURRENT FILING DATE: 2003-01-31
; PRIOR APPLICATION NUMBER: US 08/831,309
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 78
; LENGTH: 306
; TYPE: PRT

```

; ORGANISM: Helicobacter pylori
US-09-988-067B-78

Query Match 64.8%; Score 35; DB 11; Length 306;
Best Local Similarity 66.7%; Pred. No. 2.6e+02;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
|| || || ||
Db 91 CNLRNHLAC 99

RESULT 4

US-09-815-242-10452
; Sequence 10452, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10452
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10452

Query Match 64.8%; Score 35; DB 9; Length 378;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 RLHLRC 9

Db |||||
 132 RLHLRC 137

RESULT 5

US-09-815-242-5133

; Sequence 5133, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5133
; LENGTH: 387
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5133

Query Match 64.8%; Score 35; DB 9; Length 387;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 RLHLRC 9
 |||||
Db 136 RLHLRC 141

RESULT 6

US-09-954-433-8

; Sequence 8, Application US/09954433
; Patent No. US20020155562A1

; GENERAL INFORMATION:
;
; APPLICANT: Floh , Leopold
; Koenig, Kerstin
; Menge, Ulrich
;
; TITLE OF INVENTION: Glutathionylspermidine Synthetase and
; Processes for Recovery and Use Thereof
;
; NUMBER OF SEQUENCES: 9
;
; CORRESPONDENCE ADDRESS:
;
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;
; STREET: 233 South Wacker Drive/ 6300 Sears Tower
;
; CITY: Chicago
;
; STATE: Illinois
;
; COUNTRY: United States of America
;
; ZIP: 60606
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/09/954,433
;
; FILING DATE: 17-Sep-2001
;
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
;
; APPLICATION NUMBER: 09/330,740
;
; FILING DATE: <Unknown>
;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: Zeller, James P.
;
; REGISTRATION NUMBER: 28,491
;
; REFERENCE/DOCKET NUMBER: 29473/35677
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (312) 474-6300
;
; TELEFAX: (312) 474-0448
;
; INFORMATION FOR SEQ ID NO: 8:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 573 amino acids
;
; TYPE: amino acid
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
;
; HYPOTHETICAL: NO
;
; ANTI-SENSE: NO
;
; FEATURE:
;
; NAME/KEY: Modified-site
;
; LOCATION: 191
;
; OTHER INFORMATION: /note= "Xaa = Lys or Asn"
;
; FEATURE:
;
; NAME/KEY: Modified-site
;
; LOCATION: 463
;
; OTHER INFORMATION: /note= "Xaa = Val or Asp"
;
; FEATURE:
;
; NAME/KEY: Modified-site
;
; LOCATION: 479
;
; OTHER INFORMATION: /note= "Xaa = Val or Gly"
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-954-433-8

Query Match 64.8%; Score 35; DB 10; Length 573;
Best Local Similarity 55.6%; Pred. No. 4.6e+02;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
|: |||
Db 236 CDHEFHLRC 244

RESULT 7

US-09-749-601A-10
; Sequence 10, Application US/09749601A
; Patent No. US20020128460A1
; GENERAL INFORMATION:
; APPLICANT: Nicolaides, Nicholas
; APPLICANT: Grasso, Luigi
; APPLICANT: Sass, Philip
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Vogelstein, Bert
; TITLE OF INVENTION: A method for generating hypermutable
; TITLE OF INVENTION: plants
; FILE REFERENCE: 01107.00069
; CURRENT APPLICATION NUMBER: US/09/749,601A
; CURRENT FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: 60/183,333
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1151
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-749-601A-10

Query Match 63.9%; Score 34.5; DB 10; Length 1151;
Best Local Similarity 60.0%; Pred. No. 1.1e+03;
Matches 6; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Qy 1 CN-SRLHLRC 9
|| |:|:|
Db 800 CNASQMLKC 809

RESULT 8

US-09-912-697-33
; Sequence 33, Application US/09912697
; Publication No. US20030068808A1
; GENERAL INFORMATION:
; APPLICANT: Nicolaides, Nicholas C
; APPLICANT: Sass, Philip M
; APPLICANT: Grasso, Luigi M
; APPLICANT: Kline, J Bradford
; TITLE OF INVENTION: METHODS FOR GENERATING ANTIBIOTIC RESISTANT MICROBES AND
NOVEL
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: MOR-0040
; CURRENT APPLICATION NUMBER: US/09/912,697

; CURRENT FILING DATE: 2001-07-25
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 1151
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-912-697-33

Query Match 63.9%; Score 34.5; DB 11; Length 1151;
Best Local Similarity 60.0%; Pred. No. 1.1e+03;
Matches 6; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Qy 1 CN-SRLHLRC 9
|||::|||:
Db 800 CNASQMHLKC 809

RESULT 9

US-09-983-802-580
; Sequence 580, Application US/09983802
; Publication No. US20030022185A1
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: PZ010P1
; CURRENT APPLICATION NUMBER: US/09/983,802
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/227,357
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/US98/13684
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,926
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,793
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,925
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,929
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,803
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,732
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,931
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,932
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,916
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,930
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,918
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,920
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,733

```

; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,795
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,919
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,928
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,722
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,723
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,948
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,949
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,953
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,950
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,947
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,964
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/056,360
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,684
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,984
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,954
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,785
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,664
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,660
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,661
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 580
;   LENGTH: 40
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-983-802-580

```

```

Query Match          63.0%;  Score 34;  DB 11;  Length 40;
Best Local Similarity 44.4%;  Pred. No. 59;
Matches      4;  Conservative      4;  Mismatches      1;  Indels      0;  Gaps      0;

```

```

Qy      1  CNSRLHLRC 9
        | :|:|::|
Db      2  CVTRMHVKC 10

```

RESULT 10

US-10-029-386-28271
 ; Sequence 28271, Application US/10029386
 ; Publication No. US20030194704A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharron G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES
 USEFUL FOR GENE
 ; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
 ; FILE REFERENCE: AEOMICA-X-2
 ; CURRENT APPLICATION NUMBER: US/10/029,386
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 34288
 ; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 28271
 ; LENGTH: 65
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO CHR22_12.0
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.1
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
 US-10-029-386-28271

Query Match 63.0%; Score 34; DB 12; Length 65;
 Best Local Similarity 55.6%; Pred. No. 92;
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
 | | : | |
 Db 51 CTSSMHLSC 59

RESULT 11
 US-09-771-161A-118
 ; Sequence 118, Application US/09771161A
 ; Patent No. US20020110811A1
 ; GENERAL INFORMATION:
 ; APPLICANT: LEVINE, et al.
 ; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
 ; FILE REFERENCE: 802620-2005.1
 ; CURRENT APPLICATION NUMBER: US/09/771,161A
 ; CURRENT FILING DATE: 2001-01-26
 ; PRIOR APPLICATION NUMBER: 09/724,676
 ; PRIOR FILING DATE: 2000-11-28
 ; PRIOR APPLICATION NUMBER: 136776
 ; PRIOR FILING DATE: 2000-06-15
 ; PRIOR APPLICATION NUMBER: 135619
 ; PRIOR FILING DATE: 2000-04-12
 ; NUMBER OF SEQ ID NOS: 273
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 118

; LENGTH: 319
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-771-161A-118

Query Match 63.0%; Score 34; DB 10; Length 319;
Best Local Similarity 44.4%; Pred. No. 4e+02;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
||: |:| :|
Db 192 CNAAIHKKC 200

RESULT 12

US-09-771-161A-209
; Sequence 209, Application US/09771161A
; Patent No. US20020110811A1
; GENERAL INFORMATION:
; APPLICANT: LEVINE, et al.
; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
; FILE REFERENCE: 802620-2005.1
; CURRENT APPLICATION NUMBER: US/09/771,161A
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 09/724,676
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 136776
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 135619
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 273
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 209
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-771-161A-209

Query Match 63.0%; Score 34; DB 10; Length 676;
Best Local Similarity 44.4%; Pred. No. 7.9e+02;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
||: |:| :|
Db 192 CNAAIHKKC 200

RESULT 13

US-09-749-956-2
; Sequence 2, Application US/09749956
; Patent No. US20020068271A1
; GENERAL INFORMATION:
; APPLICANT: La Jolla Institute For Allergy
; APPLICANT: Altman, Amnon
; APPLICANT: Coudronniere, No. US20020068271A1wenn
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS CAPABLE OF MODULATING
PROTEIN KINASE C

; TITLE OF INVENTION: THETA (PKC?) ACTIVITY
; FILE REFERENCE: 051501/0276390
; CURRENT APPLICATION NUMBER: US/09/749,956
; CURRENT FILING DATE: 2000-12-27
; PRIOR APPLICATION NUMBER: 60/173,171
; PRIOR FILING DATE: 1999-12-27
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 706
; TYPE: PRT
; ORGANISM: Human
US-09-749-956-2

Query Match 63.0%; Score 34; DB 9; Length 706;
Best Local Similarity 44.4%; Pred. No. 8.2e+02;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
||: |:| :|
Db 193 CNAAIHKKC 201

RESULT 14

US-10-100-818-4

; Sequence 4, Application US/10100818
; Publication No. US20030176333A1
; GENERAL INFORMATION:
; APPLICANT: Lorens, James B.
; APPLICANT: Xu, Weiduan
; APPLICANT: Bogenberger, Jakob
; APPLICANT: Rigel Pharmaceuticals, Incorporated
; TITLE OF INVENTION: CASPR3: Modulators of Angiogenesis
; FILE REFERENCE: 021044-001900US
; CURRENT APPLICATION NUMBER: US/10/100,818
; CURRENT FILING DATE: 2002-03-18
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1154
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: full length contactin associated protein 3
; OTHER INFORMATION: (CASPR3)
US-10-100-818-4

Query Match 63.0%; Score 34; DB 12; Length 1154;
Best Local Similarity 66.7%; Pred. No. 1.3e+03;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
| || |||
Db 542 CEQRLALRC 550

RESULT 15

US-10-100-608B-9
; Sequence 9, Application US/10100608B
; Publication No. US20030104412A1
; GENERAL INFORMATION:
; APPLICANT: Heiskala, Marja
; TITLE OF INVENTION: REG-LIKE PROTEIN
; FILE REFERENCE: CDS-261
; CURRENT APPLICATION NUMBER: US/10/100,608B
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 60/276,414
; PRIOR FILING DATE: 2002-03-16
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Human
US-10-100-608B-9

Query Match 61.1%; Score 33; DB 15; Length 29;
Best Local Similarity 55.6%; Pred. No. 65;
Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
|| || || ||
Db 18 CNKRQHFLC 26

Search completed: November 13, 2003, 09:58:27
Job time : 19.6562 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:38:30 ; Search time 9.375 Seconds
(without alignments)
92.322 Million cell updates/sec

Title: US-09-228-866-1
Perfect score: 54
Sequence: 1 CNSRLHLRC 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_76:*
 1: pir1:*
 2: pir2:*
 3: pir3:*
 4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	% Query					Description
	No.	Score	Match	Length	ID	
1	39	72.2	1091	2	T13170	diaphanous protein
2	38	70.4	315	2	T50561	SIN1 protein [imp
3	38	70.4	514	2	A82581	periplasmic protei
4	38	70.4	781	1	TVFFDF	protein kinase Dra
5	37	68.5	464	2	T08924	hypothetical prote
6	37	68.5	736	2	S47645	tMDC I protein - c
7	37	68.5	1430	2	T27924	hypothetical prote
8	36	66.7	146	2	T02395	hypothetical prote
9	36	66.7	431	2	S55134	probable membrane
10	36	66.7	1020	2	E86165	F15K9.2 protein -
11	35.5	65.7	689	2	H97836	DNA ligase (NAD) (
12	35	64.8	13	2	JH0460	corticostatic pept
13	35	64.8	180	2	E84774	probable RING zinc
14	35	64.8	187	2	S62511	probable peptide m
15	35	64.8	296	1	S37312	transcription acti
16	35	64.8	305	2	F71968	hypothetical prote
17	35	64.8	306	1	H64539	conserved hypothet
18	35	64.8	315	2	T47971	seven in absentia-
19	35	64.8	378	2	F91263	phosphonate metabo
20	35	64.8	378	2	C86104	phosphonate metabo
21	35	64.8	378	2	S56323	hisM protein-like
22	35	64.8	387	2	H83223	conserved hypothet
23	35	64.8	426	2	G96528	protein F27J15.2 [
24	35	64.8	722	2	T26297	hypothetical prote
25	35	64.8	746	2	G84605	hypothetical prote
26	35	64.8	1941	2	T30554	ubiquitin-protein
27	35	64.8	2422	2	T12687	ALR protein homolo
28	34.5	63.9	1151	2	T04657	hypothetical prote
29	34	63.0	60	2	B69186	ribosomal protein
30	34	63.0	75	2	T44224	hypothetical prote
31	34	63.0	121	2	S66830	hypothetical prote
32	34	63.0	164	2	T24200	hypothetical prote
33	34	63.0	166	2	T18513	hypothetical prote
34	34	63.0	224	2	T17983	hypothetical prote
35	34	63.0	235	2	H72623	hypothetical prote
36	34	63.0	318	2	H69959	hypothetical prote
37	34	63.0	321	2	AE1606	B. subtilis YqhQ p
38	34	63.0	321	2	AI1243	B. subtilis YqhQ p
39	34	63.0	325	2	C84000	hypothetical prote
40	34	63.0	344	2	T05064	hypothetical prote

41	34	63.0	463	2	T01872	hypothetical prote
42	34	63.0	485	2	S34800	sulfate transport
43	34	63.0	673	1	KIRTC	protein kinase C (
44	34	63.0	674	1	KIMSCD	protein kinase C (
45	34	63.0	676	1	S35704	protein kinase C (

ALIGNMENTS

RESULT 1

T13170

diaphanous protein - fruit fly (*Drosophila melanogaster*)

C;Species: *Drosophila melanogaster*

C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 17-Nov-2000

C;Accession: T13170

R;Castrillon, D.H.; Wasserman, S.A.; Castrillon, D.H.; Wasserman, S.A.

Development 120, 3367-3377, 1994

A;Title: Diaphanous is required for cytokinesis in *Drosophila* and shares domains of similarity with the products of the limb deformity gene.

A;Reference number: Z17626; MUID:95121197; PMID:7821209

A;Accession: T13170

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-1091 <CAS>

A;Cross-references: EMBL:U11288; NID:g575926; PID:g575927; PIDN:AAA67715.1

C;Genetics:

A;Gene: dia

A;Cross-references: FlyBase:FBgn0011202

A;Map position: 2L

Query Match 72.2%; Score 39; DB 2; Length 1091;
 Best Local Similarity 87.5%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 NSRLHLRC 9
 | |||||
 Db 307 NFRHLHLRC 314

RESULT 2

T50561

SINA1 protein [imported] - *Vitis vinifera*

C;Species: *Vitis vinifera*

C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 24-May-2001

C;Accession: T50561

R;Brehm, I.; Korfei, M.; Preisig-Mueller, R.; Kindl, H.

submitted to the EMBL Data Library, November 1998

A;Description: A nuclear localized zinc finger protein found in a plant is homologous to the *Drosophila* signal transducing factor seven in absentia.

A;Reference number: Z25132

A;Accession: T50561

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-315 <BRE>

A;Cross-references: EMBL:Y18471; PIDN:CAB40577.1

C;Superfamily: *Drosophila* developmental protein sina; RING finger homology

Query Match 70.4%; Score 38; DB 2; Length 315;
Best Local Similarity 66.7%; Pred. No. 13;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
| ||:| ||
Db 78 CKSRVHNRC 86

RESULT 3

A82581

periplasmic proteinase XF2241 [imported] - *Xylella fastidiosa* (strain 9a5c)

C;Species: *Xylella fastidiosa*

C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000

C;Accession: A82581

R;anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis, Sao Paulo, Brazil.

Nature 406, 151-157, 2000

A;Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.

A;Reference number: A82515; MUID:20365717; PMID:10910347

A;Note: for a complete list of authors see reference number A59328 below

A;Accession: A82581

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-514 <SIM>

A;Cross-references: GB:AE004037; GB:AE003849; NID:g9107394; PIDN:AAF85040.1; GSPDB:GN00128; XFSC:XF2241

A;Experimental source: strain 9a5c

R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Alves, L.M.C.; Araya, J.E.; Baia, G.S.; Baptista, C.S.; Barros, M.H.; Bonaccorsi, E.D.; Bordin, S.; Bove, J.M.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.; Colauto, N.B.; Colombo, C.; Costa, F.F.; Costa, M.C.R.; Costa-Neto, C.M.; Coutinho, L.L.; Cristofani, M.; Dias-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme, M.; Furlan, L.R.; Garnier, M.; Goldman, G.H.; Goldman, M.H.S.; Gomes, S.L.; Gruber, A.; Ho, P.L.; Hoheisel, J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigret, F.; Lambais, M.R.; Leite, L.C.C.; Lemos, E.G.M.; Lemos, M.V.F.; Lopes, S.A.; Lopes, C.R.; Machado, J.A.; Machado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.A.L.

A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; Monteiro-Vitorello, C.B.; Moon, D.H.; Nagai, M.A.; Nascimento, A.L.T.O.; Netto, L.E.S.; Nhani Jr., A.; Nobrega, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Paris, A.; Peixoto, B.R.; Pereira, G.A.G.; Pereira Jr., H.A.; Pesquero, J.B.; Quaggio, R.B.; Roberto, P.G.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki, H.E.

A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira, J.F.; Silvestri, M.L.Z.; Siqueira, W.J.; de Souza, A.A.; de Souza, A.P.; Terenzi, M.F.; Truffi, D.; Tsai, S.M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Zago, M.A.; Zatz, M.; Meidanis, J.; Setubal, J.C.

A;Reference number: A59328

A;Contents: annotation
C;Genetics:
A;Gene: XF2241
C;Superfamily: Helicobacter serine proteinase

Query Match 70.4%; Score 38; DB 2; Length 514;
Best Local Similarity 75.0%; Pred. No. 19;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 NSRLHLRC 9
|||:| ||
Db 2 NSRIHTRC 9

RESULT 4

TVFFDF

protein kinase Draf-1 (EC 2.7.1.-) - fruit fly (*Drosophila melanogaster*)
N;Alternate names: Draf-1 proto-oncogene protein-serine/threonine kinase;
kinase-related transforming protein Draf-1; pole-hole protein

C;Species: *Drosophila melanogaster*

C;Date: 31-Mar-1991 #sequence_revision 23-Feb-1996 #text_change 23-Feb-1997

C;Accession: S00393; S60191; A27808; S33602

R;Nishida, Y.; Hata, M.; Ayaki, T.; Ryo, H.; Yamagata, M.; Shimizu, K.;
Nishizuka, Y.

EMBO J. 7, 775-781, 1988

A;Title: Proliferation of both somatic and germ cells is affected in the
Drosophila mutants of raf proto-oncogene.

A;Reference number: S00393; MUID:88283647; PMID:3135183

A;Accession: S00393

A;Molecule type: DNA

A;Residues: 1-781 <NIS>

A;Cross-references: EMBL:X07181

A;Note: the assignment of the start codon has been revised in reference S33602

A;Accession: S60191

A;Molecule type: mRNA

A;Residues: 148-781 <NIS2>

R;Mark, G.E.; MacIntyre, R.J.; Digan, M.E.; Ambrosio, L.; Perrimon, N.
Mol. Cell. Biol. 7, 2134-2140, 1987

A;Title: *Drosophila melanogaster* homologs of the raf oncogene.

A;Reference number: A27808; MUID:87257926; PMID:3037346

A;Accession: A27808

A;Molecule type: mRNA

A;Residues: 'LQ',465-519,'R',521,'A',523-570,'R',572-699,'PQAL',704-
713,'PT',716-753 <MAR>

R;Sprenger, F.; Trosclair, M.M.; Morrison, D.K.
Mol. Cell. Biol. 13, 1163-1172, 1993

A;Title: Biochemical analysis of torso and D-raf during *Drosophila*
embryogenesis: implications for terminal signal transduction.

A;Reference number: S33602; MUID:93140754; PMID:8423783

A;Contents: annotation

A;Note: this is a revision of the assignment of the start codon in reference
S00393

A;Note: the authors call the N-terminal extended version of the protein Draf-3

A;Note: the cited sequence in S33602 shows Pro for residue 342

C;Genetics:

A;Gene: Draf-1

A;Cross-references: FlyBase:FBgn0003079

A;Map position: X 2F
A;Introns: 417/3; 464/3; 589/2
C;Superfamily: protein kinase A-raf; protein kinase C zinc-binding repeat
homology; protein kinase homology
C;Keywords: ATP; phosphotransferase; proto-oncogene; serine/threonine-specific
protein kinase; transforming protein
F;265-310/Domain: protein kinase C zinc-binding repeat homology <KZ2>
F;469-735/Domain: protein kinase homology <KIN>
F;477-485/Region: protein kinase ATP-binding motif
F;497/Active site: Lys #status predicted

Query Match 70.4%; Score 38; DB 1; Length 781;
Best Local Similarity 66.7%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CNSRLHLRC 9
|||
Db 294 CNFRFHQRC 302

RESULT 5

T08924

hypothetical protein T15N24.30 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 02-Sep-2000

C;Accession: T08924

R;Bevan, M.; Zimmermann, W.; Grueneisen, A.; Wambutt, R.; Bancroft, I.; Mewes,
H.W.; Mayer, K.F.X.; Lemcke, K.; Schueller, C.

submitted to the Protein Sequence Database, May 1999

A;Reference number: Z16518

A;Accession: T08924

A;Molecule type: DNA

A;Residues: 1-464 <BEV>

A;Cross-references: EMBL:AL078465; GSPDB:GN00062; ATSP:T15N24.30

A;Experimental source: cultivar Columbia; BAC clone T15N24

C;Genetics:

A;Gene: ATSP:T15N24.30

A;Map position: 4

A;Introns: 38/2; 84/3; 106/3; 297/2; 416/3

C;Superfamily: RING finger homology

F;414-464/Domain: RING finger homology <RRN>

Query Match 68.5%; Score 37; DB 2; Length 464;
Best Local Similarity 55.6%; Pred. No. 27;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CNSRLHLRC 9
|:|
Db 436 CSHRFHLKC 444

RESULT 6

S47645

tMDC I protein - crab-eating macaque

C;Species: Macaca fascicularis (crab-eating macaque)

C;Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 21-Jul-2000

C;Accession: S47645

R;Barker, H.L.; Perry, A.C.F.; Jones, R.; Hall, L.
 Biochim. Biophys. Acta 1218, 429-431, 1994
 A;Title: Sequence and expression of a monkey testicular transcript encoding tMDC
 I, a novel member of the metalloproteinase-like, disintegrin-like, cysteine-rich
 (MDC) protein family.
 A;Reference number: S47645; MUID:94325353; PMID:8049267
 A;Accession: S47645
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-736 <BAR>
 A;Cross-references: EMBL:X76637; NID:g535016; PIDN:CAA54085.1; PID:g535017
 C;Superfamily: mouse meltrin alpha; disintegrin homology
 F;392-477/Domain: disintegrin homology <DIS>

Query Match 68.5%; Score 37; DB 2; Length 736;
 Best Local Similarity 55.6%; Pred. No. 41;
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
 || || :|
 Db 629 CNDRFHCQC 637

RESULT 7

T27924
 hypothetical protein ZK593.4 - *Caenorhabditis elegans*
 C;Species: *Caenorhabditis elegans*
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
 C;Accession: T27924
 R;McMurray, A.
 submitted to the EMBL Data Library, February 1996
 A;Reference number: Z20440
 A;Accession: T27924
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-1430 <WIL>
 A;Cross-references: EMBL:Z69385; PIDN:CAA93426.1; GSPDB:GN00022; CESP:ZK593.4
 A;Experimental source: clone ZK593
 C;Genetics:
 A;Gene: CESP:ZK593.4
 A;Map position: 4
 A;Introns: 48/3; 92/3; 238/2; 254/1; 924/2; 987/1; 1085/3; 1304/2; 1404/1
 C;Superfamily: human retinoblastoma binding protein 2

Query Match 68.5%; Score 37; DB 2; Length 1430;
 Best Local Similarity 55.6%; Pred. No. 73;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
 |:| |:|
 Db 1179 CDSEFHVRC 1187

RESULT 8

T02395
 hypothetical protein At2g44400 [imported] - *Arabidopsis thaliana*
 N;Alternate names: hypothetical protein F4I1.21

C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 02-Feb-2001
 C;Accession: T02395; A84878
 R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.
 submitted to the EMBL Data Library, May 1998
 A;Description: Arabidopsis thaliana chromosome II BAC F4I1 genomic sequence.
 A;Reference number: Z14667
 A;Accession: T02395
 A;Status: translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-146 <ROU>
 A;Cross-references: EMBL:AC004521; NID:g3128166; PID:g3128182
 A;Experimental source: cultivar Columbia
 R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Mason, T.M.; Bowman, C.L.; Barnstead, M.E.; Feldblyum, T.V.; Buell, C.R.; Ketchum, K.A.; Lee, J.J.; Ronning, C.M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J.; Gill, J.E.; Adams, M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.; Copenhagen, G.P.; Preuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.C.
 Nature 402, 761-768, 1999
 A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A;Reference number: A84420; MUID:20083487; PMID:10617197
 A;Accession: A84878
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-146 <STO>
 A;Cross-references: GB:AE002093; NID:g3128182; PIDN:AAC16086.1; GSPDB:GN00139
 C;Genetics:
 A;Gene: At2g44400; F4I1.21
 A;Map position: 2

Query Match 66.7%; Score 36; DB 2; Length 146;
 Best Local Similarity 55.6%; Pred. No. 15;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
 || :||:|
 Db 18 CNFYIHLKC 26

RESULT 9
 S55134
 probable membrane protein YMR187c - yeast (Saccharomyces cerevisiae)
 N;Alternate names: hypothetical protein YM8010.17c
 C;Species: Saccharomyces cerevisiae
 C;Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
 C;Accession: S55134
 R;Churcher, C.M.
 submitted to the EMBL Data Library, June 1995
 A;Reference number: S55118
 A;Accession: S55134
 A;Molecule type: DNA
 A;Residues: 1-431 <CHU>

A;Cross-references: EMBL:Z49808; NID:g854440; PID:g854457; GSPDB:GN00013;
MIPS:YMR187c
A;Experimental source: strain AB972
C;Genetics:
A;Gene: MIPS:YMR187c
A;Cross-references: SGD:S0004799
A;Map position: 13R
C;Superfamily: Saccharomyces cerevisiae probable membrane protein YMR187c
C;Keywords: transmembrane protein
F;228-244/Domain: transmembrane #status predicted <TMM>

Query Match 66.7%; Score 36; DB 2; Length 431;
Best Local Similarity 55.6%; Pred. No. 39;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
|| ::||
Db 37 CNLQIHKRC 45

RESULT 10

E86165

F15K9.2 protein - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001

C;Accession: E86165

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.;
Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.;
Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.;
Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.;
Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.;
Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.;
Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-
Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu,
S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda,
M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.;
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.;
Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken,
S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.;
Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: E86165

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1020 <STO>

A;Cross-references: GB:AE005172; NID:g3850588; PIDN:AAC72128.1; GSPDB:GN00141

C;Genetics:

A;Map position: 1

Query Match 66.7%; Score 36; DB 2; Length 1020;
Best Local Similarity 75.0%; Pred. No. 83;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CNSRLHLR 8
 | :| | | |
 Db 645 CQSKLHLR 652

RESULT 11

H97836

DNA ligase (NAD) (EC 6.5.1.2) - Rickettsia conorii (strain Malish 7)

C;Species: Rickettsia conorii

C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 03-Jun-2002

C;Accession: H97836

R;Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.;
 Samson, D.; Roux, V.; Cossart, P.; Weissenbach, J.; Claverie, J.M.; Raoult, D.
 Science 293, 2093-2098, 2001

A;Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia
 prowazekii.

A;Reference number: A97700; MUID:21442074; PMID:11557893

A;Accession: H97836

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-689 <KUR>

A;Cross-references: GB:AE006914; PIDN:AAL03634.1; PID:g15620219; GSPDB:GN00173

C;Genetics:

A;Gene: lig

C;Superfamily: polydeoxyribonucleotide synthase (NAD+)

C;Keywords: ligase

Query Match 65.7%; Score 35.5; DB 2; Length 689;
 Best Local Similarity 43.8%; Pred. No. 73;
 Matches 7; Conservative 2; Mismatches 0; Indels 7; Gaps 1;

Qy 1 CNSRLH-----LRC 9
 |||:| | :| |
 Db 416 CNSKLHYTPEDIIRC 431

RESULT 12

JH0460

corticostatic peptide GP-CS3 - guinea pig

C;Species: Cavia porcellus (guinea pig)

C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 18-Aug-2000

C;Accession: JH0460

R;Hu, J.; Bennett, H.P.J.; Lazure, C.; Solomon, S.

Biochem. Biophys. Res. Commun. 180, 558-565, 1991

A;Title: Isolation and characterization of corticostatic peptides from guinea
 pig bone marrow.

A;Reference number: JH0458; MUID:92062075; PMID:1659400

A;Accession: JH0460

A;Molecule type: protein

A;Residues: 1-13 <HUJ>

A;Experimental source: bone marrow

A;Note: this is a dimer having an antiparallel configuration

C;Comment: This peptide belongs to a family of Cys-rich, cationic peptides of
 low molecular weight.

C;Comment: This peptide has antimicrobial activity by a non-oxygen-dependent
 mechanism.

C;Superfamily: unassigned animal peptides

F;5/Disulfide bonds: interchain (to 13) #status experimental
F;7/Disulfide bonds: interchain (to 11) #status experimental
F;11/Disulfide bonds: interchain (to 7) #status experimental
F;13/Disulfide bonds: interchain (to 5) #status experimental

Query Match 64.8%; Score 35; DB 2; Length 13;
Best Local Similarity 66.7%; Pred. No. 2.8;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
| ||| ||
Db 5 CFCRLHCRC 13

RESULT 13

E84774

probable RING zinc finger protein [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C;Accession: E84774

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.;
Fujii, C.Y.; Mason, T.M.; Bowman, C.L.; Barnstead, M.E.; Feldblyum, T.V.; Buell,
C.R.; Ketchum, K.A.; Lee, J.J.; Ronning, C.M.; Koo, H.; Moffat, K.S.; Cronin,
L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J.; Gill, J.E.; Adams,
M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.; Copenhaver,
G.P.; Preuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser,
C.M.; Venter, J.C.

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: E84774

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-180 <STO>

A;Cross-references: GB:AE002093; NID:g4510378; PIDN:AAD21466.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g35910

A;Map position: 2

Query Match 64.8%; Score 35; DB 2; Length 180;
Best Local Similarity 55.6%; Pred. No. 28;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
|| ||:
Db 129 CNHLFHLKC 137

RESULT 14

S62511

probable peptide methionine sulfoxide reductase - fission yeast
(Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe

C;Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 10-Dec-1999

C;Accession: T38506; S62511

R;Jones, L.; Murphy, L.; McNeil, A.; Simpson, I.; Harris, D.; Barrell, B.G.;
Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, October 1995
A;Reference number: Z21798
A;Accession: T38506
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-187 <JO2>
A;Cross-references: EMBL:Z66525; NID:g1044926; PIDN:CAA91427.1; PID:g1044931;
GSPDB:GN00066; SPDB:SPAC29E6.05c
A;Experimental source: strain 972h-; cosmid c29E6
C;Genetics:
A;Gene: SPDB:SPAC29E6.05c
A;Map position: 1
C;Superfamily: peptide methionine sulfoxide reductase

Query Match 64.8%; Score 35; DB 2; Length 187;
Best Local Similarity 44.4%; Pred. No. 29;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
|:|:|:|:
Db 159 CSSRMNIKC 167

RESULT 15

S37312

transcription activator hlyT NhaR VC0677 [similarity] - *Vibrio cholerae* (strain N16961 serogroup O1)

C;Species: *Vibrio cholerae*

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 02-Feb-2001

C;Accession: S37312; G82292

R;Williams, S.G.; Attridge, S.R.; Manning, P.A.

Mol. Microbiol. 9, 751-760, 1993

A;Title: The transcriptional activator HlyU of *Vibrio cholerae*: nucleotide sequence and role in virulence gene expression.

A;Reference number: S37312; MUID:94049116; PMID:8231807

A;Accession: S37312

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-296 <WIL>

A;Cross-references: EMBL:X66866; NID:g403330; PID:g403331

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1992

R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.;
Dodson, R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Umayam, L.A.; Gill,
S.R.; Nelson, K.E.; Read, T.D.; Tettelin, H.; Richardson, D.; Ermolaeva, M.D.;
Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.; McDonald, L.;
Utterback, T.; Fleishmann, R.D.; Nierman, W.C.; White, O.; Salzberg, S.L.;
Smith, H.O.; Colwell, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000

A;Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.

A;Reference number: A82035; MUID:20406833; PMID:10952301

A;Accession: G82292

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-296 <HEI>
A;Cross-references: GB:AE004154; GB:AE003852; NID:g9655115; PIDN:AAF93842.1;
GSPDB:GN00126; TIGR:VC0677
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: hlyT; VC0677
A;Map position: 1
C;Superfamily: regulatory protein nhaR
C;Keywords: DNA binding; transcription regulation
F;21-40/Region: helix-turn-helix motif

Query Match 64.8%; Score 35; DB 1; Length 296;
Best Local Similarity 62.5%; Pred. No. 43;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 NSRLHLRC 9
::|:||||
Db 119 DNRIHLRC 126

Search completed: November 13, 2003, 09:52:49
Job time : 11.375 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 5.15625 Seconds
(without alignments)
82.083 Million cell updates/sec

Title: US-09-228-866-1
Perfect score: 54
Sequence: 1 CNSRLHLRC 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	42	77.8	302	1	AS14_HUMAN	Q8wxk2 homo sapien
2	39	72.2	1091	1	DIA_DROME	P48608 drosophila
3	38	70.4	781	1	KRAF_DROME	P11346 drosophila
4	36	66.7	431	1	YM49_YEAST	Q03236 saccharomyc
5	36	66.7	1342	1	Z335_HUMAN	Q9h4z2 homo sapien
6	36	66.7	7073	1	R1AB_CVHSA	P59641 h replicase
7	35	64.8	296	1	NHAR_VIBCH	P52692 vibrio chol
8	35	64.8	378	1	PHNM_ECOLI	P16689 escherichia
9	35	64.8	652	1	TRYS_CRIFA	O60993 crithidia f
10	35	64.8	1941	1	UBR1_KLULA	O60014 kluyveromyc
11	34	63.0	60	1	RL37_METTH	O26744 methanobact
12	34	63.0	235	1	ECR1_AERPE	Q9yc02 aeropyrum p
13	34	63.0	318	1	YQHQ_BACSU	P54515 bacillus su
14	34	63.0	433	1	AS14_MOUSE	Q8vhs7 mus musculu
15	34	63.0	485	1	NO70_SOYBN	Q02920 glycine max
16	34	63.0	673	1	KPCD_RAT	P09215 rattus norv
17	34	63.0	674	1	KPCD_MOUSE	P28867 mus musculu
18	34	63.0	676	1	KPCD_HUMAN	Q05655 homo sapien
19	34	63.0	706	1	KPCT_HUMAN	Q04759 homo sapien
20	34	63.0	707	1	KPCT_MOUSE	Q02111 mus musculu
21	34	63.0	1238	1	YQO9_CAEEL	Q09298 caenorhabdi
22	34	63.0	1288	1	CTA3_HUMAN	Q9bz76 homo sapien
23	34	63.0	2652	1	RRPB_IBVB	P26314 avian infec
24	33	61.1	530	1	RAG2_BRARE	O13034 brachydanio
25	33	61.1	533	1	RAG2_ONCMY	Q91193 oncorhynchu
26	33	61.1	596	1	LFC_MOUSE	Q60875 mus musculu
27	33	61.1	682	1	KPCL_HUMAN	P24723 homo sapien
28	33	61.1	683	1	KPCL_MOUSE	P23298 mus musculu
29	33	61.1	683	1	KPCL_RAT	Q64617 rattus norv
30	33	61.1	768	1	SC23_YEAST	P15303 saccharomyc
31	33	61.1	893	1	ARH2_HUMAN	Q92974 homo sapien
32	33	61.1	1163	1	LEPR_MACMU	Q9myl0 macaca mula
33	32	59.3	89	1	R14B_LISMO	Q8y620 listeria mo
34	32	59.3	101	1	RS14_CHLMU	Q9pld4 chlamydia m
35	32	59.3	101	1	RS14_CHLTR	O84792 chlamydia t
36	32	59.3	411	1	HEMZ_XENLA	O57478 xenopus lae
37	32	59.3	411	1	RAPS_HUMAN	Q13702 homo sapien
38	32	59.3	411	1	RAPS_MOUSE	P12672 mus musculu
39	32	59.3	458	1	EX7L_BACHD	Q9k967 bacillus ha
40	32	59.3	470	1	A1AA_ORYLA	Q91175 oryzias lat
41	32	59.3	623	1	ZP1_MOUSE	Q62005 mus musculu
42	32	59.3	697	1	AD26_MOUSE	Q9r158 mus musculu
43	32	59.3	726	1	AD20_HUMAN	O43506 homo sapien
44	32	59.3	735	1	AD02_CAVPO	Q60411 cavia porce
45	32	59.3	735	1	AD02_HUMAN	Q99965 homo sapien

ALIGNMENTS

RESULT 1
AS14_HUMAN

ID AS14_HUMAN STANDARD; PRT; 302 AA.
AC Q8WXK2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ankyrin repeat and SOCS box containing protein 14 (ASB-14).
GN ASB14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kile B.T., Nicola N.A.;
RT "SOCS box proteins."
RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: Contains 4 ANK repeats.
CC -!- SIMILARITY: Contains 1 SOCS box domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF403032; AAL57351.1; -.
DR Genew; HGNC:19766; ASB14.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001496; SOCS.
DR Pfam; PF00023; ank; 3.
DR SMART; SM00248; ANK; 4.
DR PROSITE; PS50088; ANK_REPEAT; 2.
DR PROSITE; PS50297; ANK_REP_REGION; 1.
DR PROSITE; PS50225; SOCS; 1.
KW ANK repeat; Repeat.
FT REPEAT 28 57 ANK 1.
FT REPEAT 70 99 ANK 2.
FT REPEAT 100 129 ANK 3.
FT REPEAT 131 164 ANK 4.
FT DOMAIN 236 291 SOCS BOX.
SQ SEQUENCE 302 AA; 34562 MW; 0B8C6E7219E9EF7B CRC64;

Query Match 77.8%; Score 42; DB 1; Length 302;
Best Local Similarity 77.8%; Pred. No. 0.76;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
| |||||
Db 260 CMGRHLHLRC 268

RESULT 2
DIA_DROME
ID DIA_DROME STANDARD; PRT; 1091 AA.
AC P48608; Q9VIJ7;

DT 01-FEB-1996 (Rel. 33, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Diaphanous protein.
 GN DIA OR CG1768.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95121197; PubMed=7821209;
 RA Castrillon D.H., Wasserman S.A.;
 RT "Diaphanous is required for cytokinesis in Drosophila and shares
 RT domains of similarity with the products of the limb deformity gene.";
 RL Development 120:3367-3377(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [3]
 RP FUNCTION.
 RX MEDLINE=20214846; PubMed=10751177;
 RA Afshar K., Stuart B., Wasserman S.A.;
 RT "Functional analysis of the *Drosophila* diaphanous FH protein in early
 RT embryonic development.";
 RL Development 127:1887-1897(2000).
 CC -!- FUNCTION: REQUIRED FOR CYTOKINESIS IN BOTH MITOSIS AND MEIOSIS.
 CC HAS A ROLE IN ACTIN CYTOSKELETON ORGANIZATION AND IS ESSENTIAL FOR
 CC MANY, IF NOT ALL, ACTIN-MEDIATED EVENTS INVOLVING MEMBRANE
 CC INVAGINATION. MAY SERVE AS A MEDIATOR BETWEEN SIGNALING MOLECULES
 CC AND ACTIN ORGANIZERS AT SPECIFIC PHASES OF THE CELL CYCLE.
 CC POSSIBLE COMPONENT OF THE CONTRACTILE RING OR MAY CONTROL ITS
 CC FUNCTION.
 CC -!- SUBCELLULAR LOCATION: LOCALIZES TO THE SITE WHERE THE METAPHASE
 CC FURROW IS ANTICIPATED TO FORM, TO THE GROWING TIP OF
 CC CELLULARIZATION FURROWS, AND TO CONTRACTILE RINGS.
 CC -!- DOMAIN: DRFS ARE REGULATED BY INTRAMOLECULAR GBD-DAD BINDING WHERE
 CC RHO-GTP ACTIVATES THE DRFS BY DISRUPTING THE GBD-DAD INTERACTION
 CC (BY SIMILARITY).
 CC -!- SIMILARITY: Contains 1 GTPase-binding (GBD) domain.
 CC -!- SIMILARITY: Contains 1 Formin homology 1 (FH1) domain.
 CC -!- SIMILARITY: Contains 1 Formin homology 2 (FH2) domain.
 CC -!- SIMILARITY: Contains 1 Formin homology 3 (FH3) domain.
 CC -!- SIMILARITY: Contains 1 DRF autoregulatory (DAD) domain.
 CC -!- SIMILARITY: BELONGS TO THE FORMIN HOMOLOGY FAMILY. DIAPHANOUS
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U11288; AAA67715.1; -.
 DR EMBL; AE003668; AAF53922.1; -.
 DR PIR; T13170; T13170.
 DR FlyBase; FBgn0011202; dia.
 DR InterPro; IPR003104; FH2.
 DR Pfam; PF02181; FH2; 1.
 DR SMART; SM00498; FH2; 1.
 KW Cell division; Coiled coil.
 FT DOMAIN 47 242 GBD.
 FT DOMAIN 143 448 FH3.
 FT DOMAIN 446 500 COILED COIL (POTENTIAL).
 FT DOMAIN 512 596 FH1 (PRO-RICH).
 FT DOMAIN 601 1044 FH2.
 FT DOMAIN 967 1021 COILED COIL (POTENTIAL).
 FT DOMAIN 1027 1041 DAD.
 FT DOMAIN 1050 1053 ARG-RICH (BASIC).
 FT DOMAIN 512 518 POLY-PRO.
 FT DOMAIN 519 522 POLY-GLY.
 FT DOMAIN 524 532 POLY-PRO.

FT DOMAIN 539 548 POLY-PRO.
 FT DOMAIN 554 561 POLY-PRO.
 FT DOMAIN 566 572 POLY-PRO.
 FT DOMAIN 581 585 POLY-PRO.
 FT CONFLICT 733 733 H -> Q (IN REF. 1).
 SQ SEQUENCE 1091 AA; 123170 MW; A4379D7A089B5EE7 CRC64;

Query Match 72.2%; Score 39; DB 1; Length 1091;
 Best Local Similarity 87.5%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 NSRLHLRC 9
 | |||||
 Db 307 NFRLHLRC 314

RESULT 3

KRAF_DROME

ID KRAF_DROME STANDARD; PRT; 781 AA.
 AC P11346;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE RAF homolog serine/threonine-protein kinase dRAF-1 (EC 2.7.1.-)
 DE (Pole-hole protein).
 GN PHL OR DRAF-1 OR D-RAF.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88283647; PubMed=3135183;
 RA Nishida Y., Hata M., Ayaki T., Ryo H., Yamagata M., Shimizu K.,
 RA Nishizuka Y.;
 RT "Proliferation of both somatic and germ cells is affected in the
 RT Drosophila mutants of raf proto-oncogene."
 RL EMBO J. 7:775-781(1988).
 RN [2]
 RP SEQUENCE OF 465-753 FROM N.A.
 RX MEDLINE=87257926; PubMed=3037346;
 RA Mark G.E., Macintyre R.J., Digan M.E., Ambrosio L., Perrimon N.;
 RT "Drosophila melanogaster homologs of the raf oncogene."
 RL Mol. Cell. Biol. 7:2134-2140(1987).
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=93140754; PubMed=8423783;
 RA Sprenger F., Torsocclair M.M., Morrison D.K.;
 RT "Biochemical analysis of torso and D-raf during Drosophila
 RT embryogenesis: implications for terminal signal transduction."
 RL Mol. Cell. Biol. 13:1163-1172(1993).
 CC -!- FUNCTION: SERINE/THREONINE KINASE REQUIRED IN THE EARLY EMBRYO
 CC FOR THE FORMATION OF TERMINAL STRUCTURE. ALSO REQUIRED DURING
 CC THE PROLIFERATION OF IMAGINAL CELLS. MAY ACT DOWNSTREAM OF RAS1
 CC IN THE SEV SIGNAL TRANSDUCTION PATHWAY.
 CC -!- PTM: EXTENSIVELY PHOSPHORYLATED AT 1 TO 2 H AFTER EGG LAYING.

```

CC  -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC  MIL/RAF SUBFAMILY.
CC  -!- SIMILARITY: Contains 1 zinc-dependent phorbol-ester and DAG
CC  binding domain.
CC  -!- SIMILARITY: Contains 1 Ras-binding (RBD) domain.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; X07181; CAA30166.1; ALT_INIT.
DR  EMBL; M16598; -; NOT_ANNOTATED_CDS.
DR  HSSP; P04049; 1RFA.
DR  FlyBase; FBgn0003079; ph1.
DR  GO; GO:0008069; P:dorsal/ventral axis determination, follicul. . .; IMP.
DR  GO; GO:0007369; P:gastrulation; NAS.
DR  GO; GO:0007283; P:spermatogenesis; IMP.
DR  GO; GO:0007362; P:terminal region determination; IMP.
DR  InterPro; IPR002219; DAG_PE-bind.
DR  InterPro; IPR000719; Prot_kinase.
DR  InterPro; IPR003116; RBD.
DR  InterPro; IPR002290; Ser_thr_pkinase.
DR  Pfam; PF00130; DAG_PE-bind; 1.
DR  Pfam; PF00069; pkinase; 1.
DR  Pfam; PF02196; RBD; 1.
DR  ProDom; PD0000001; Prot_kinase; 1.
DR  SMART; SM00109; C1; 1.
DR  SMART; SM00455; RBD; 1.
DR  PROSITE; PS00479; DAG_PE_BIND_DOM_1; 1.
DR  PROSITE; PS00081; DAG_PE_BIND_DOM_2; 1.
DR  PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR  PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR  PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR  PROSITE; PS00898; RBD; 1.
KW  Transferase; Serine/threonine-protein kinase; ATP-binding; Zinc;
KW  Phorbol-ester binding; Phosphorylation.
FT  DOMAIN      183    254      RAS-BINDING.
FT  DOMAIN      265    310      PHORBOL-ESTER AND DAG BINDING.
FT  DOMAIN      471    732      PROTEIN KINASE.
FT  NP_BIND     477    485      ATP (BY SIMILARITY).
FT  BINDING     497    497      ATP (BY SIMILARITY).
FT  ACT_SITE    590    590      BY SIMILARITY.
FT  CONFLICT    495    495      P -> A (IN REF. 2).
FT  CONFLICT    520    522      KKT -> RKA (IN REF. 2).
FT  CONFLICT    571    571      G -> R (IN REF. 2).
FT  CONFLICT    700    703      RRHS -> PQAL (IN REF. 2).
SQ  SEQUENCE    781 AA;  88794 MW;  DEAD54762249EADC CRC64;

```

```

Query Match          70.4%;  Score 38;  DB 1;  Length 781;
Best Local Similarity 66.7%;  Pred. No. 12;
Matches      6;  Conservative      0;  Mismatches      3;  Indels      0;  Gaps      0;

```


Db

||| |||
294 CNFRFHQRC 302

RESULT 4

YM49_YEAST

ID YM49_YEAST STANDARD; PRT; 431 AA.
AC Q03236;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical 50.3 kDa protein in HSC82-GCV2 intergenic region.
GN YMR187C OR YM8010.17C.
OS *Saccharomyces cerevisiae* (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX PubMed=9169872;
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome
RT XIII.";
RL Nature 387:90-93(1997).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z49808; CAA89920.1; -.
DR PIR; S55134; S55134.
DR SGD; S0004799; YMR187C.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 228 248 POTENTIAL.
FT TRANSMEM 279 299 POTENTIAL.
FT TRANSMEM 349 369 POTENTIAL.
FT TRANSMEM 388 408 POTENTIAL.
SQ SEQUENCE 431 AA; 50287 MW; 61165A68455B92F1 CRC64;

Query Match 66.7%; Score 36; DB 1; Length 431;
Best Local Similarity 55.6%; Pred. No. 15;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
|| ::| ||
Db 37 CNLQIHKRC 45

RESULT 5

Z335_HUMAN

ID Z335_HUMAN STANDARD; PRT; 1342 AA.

AC Q9H4Z2; Q9H684;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Zinc finger protein 335.

GN ZNF335.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21638749; PubMed=11780052;

RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,

RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,

RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,

RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,

RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,

RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,

RA Clegg S., Copley V.E., Collier R.E., Connor R.E., Corby N.R.,

RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,

RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,

RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,

RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,

RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,

RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,

RA Lehvaeslaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,

RA Marsh V.L., Martin S.L., McConnell L.J., McLay K., McMurray A.A.,

RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,

RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,

RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,

RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,

RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,

RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,

RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,

RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,

RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,

RA Rogers J.;

RT "The DNA sequence and comparative analysis of human chromosome 20.;"

RL Nature 414:865-871(2001).

RN [2]

RP SEQUENCE OF 455-1342 FROM N.A.

RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,

RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,

RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;

RT "NEDO human cDNA sequencing project.;"

RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.

CC -!- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.

CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AL162458; CAC10457.1; -.
DR EMBL; AK026157; BAB15379.1; ALT_INIT.
DR Genew; HGNC:15807; ZNF335.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 13.
DR SMART; SM00355; Znf_C2H2; 13.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 6.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 13.
KW Transcription regulation; Zinc-finger; Metal-binding; Nuclear protein;
KW DNA-binding; Repeat.
FT ZN_FING 245 268 C2H2-TYPE.
FT ZN_FING 465 487 C2H2-TYPE.
FT ZN_FING 495 517 C2H2-TYPE.
FT ZN_FING 523 545 C2H2-TYPE.
FT ZN_FING 562 584 C2H2-TYPE.
FT ZN_FING 590 612 C2H2-TYPE.
FT ZN_FING 621 643 C2H2-TYPE.
FT ZN_FING 649 672 C2H2-TYPE.
FT ZN_FING 678 701 C2H2-TYPE.
FT ZN_FING 1019 1041 C2H2-TYPE.
FT ZN_FING 1047 1069 C2H2-TYPE.
FT ZN_FING 1075 1097 C2H2-TYPE.
FT ZN_FING 1103 1126 C2H2-TYPE.
FT DOMAIN 1178 1330 GLN-RICH.
SQ SEQUENCE 1342 AA; 144892 MW; 6D230DEE0B3AE670 CRC64;

Query Match 66.7%; Score 36; DB 1; Length 1342;
Best Local Similarity 75.0%; Pred. No. 50;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 NSRLHLRC 9
| |||:|
Db 692 NLRLHVRC 699

RESULT 6

R1AB_CVHSA

ID R1AB_CVHSA STANDARD; PRT; 7073 AA.
AC P59641;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Replicase polyprotein 1ab (pp1ab) (ORF1AB) [Includes: Replicase
DE polyprotein 1a (pp1a) (ORF1A)] [Contains: Leader protein; p65 homolog;
DE Papain-like proteinase (EC 3.4.24.-) (NSP1); 3C-like proteinase
DE (EC 3.4.24.-) (3CL-PRO) (NSP2); HD2 (NSP3); NSP4; NSP5; NSP6; Growth
DE factor-like (NSP7); RNA-directed RNA polymerase (EC 2.7.7.48) (RdRp)
DE (NSP9); Helicase (Hel) (NSP10); NSP11; NSP12; NSP13].
OS Human coronavirus (strain SARS) (HCoV-SARS).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxID=227859;
RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=Isolate Urbani;
 RA Bellini W.J., Campagnoli R.P., Icenogle J.P., Monroe S.S., Nix W.A.,
 RA Oberste M.S., Pallansch M.A., Rota P.A.;
 RL Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Isolate Tor2;
 RA Marra M., Jones S.J.M., Holt R.;
 RT "The complete genome of the SARS associated coronavirus.";
 RL Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Isolate CUHK-W1;
 RA Tsui S.K.W., Lo D.Y.M., Tam J.S., Fung K.P., Chim S.S.C., Au C.C.,
 RA Chan A.H., Wan A.W.K., Au K.W., Chan C.W., Kou C.Y.C., Lam H.M.,
 RA Lam W.Y., Lau S.K., Lau Y.L., Lau Y.M., Law S.L., Law T.W., Li M.L.Y.,
 RA Tse C.H., Wong C.H., Yiu W.H., Lee C.Y., Chan A.K.C., Chiu R.W.K.,
 RA Ng E.K.O., Tong Y.K., Chan P.K.S., Au-Yeung C., Cheung J.K.L., Chu I.,
 RA Hung E.C.W., Waye M.M.Y.;
 RT "DNA sequence of a human coronavirus (CUHK-W1) from a patient with
 RT severe acute respiratory syndrome (SARS) in Hong Kong.";
 RL Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Isolate HKU-39849;
 RA Leung F.C., Zeng F., Chan C.W.M., Chan C.M.Y., Chen J., Chow K.Y.C.,
 RA Hon C.C.C., Hui R.K.H., Li J., Li V.Y.Y., Wang Y.Y., Peiris J.S.M.,
 RA Poon L.L.M.;
 RL Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
 RN [5]
 RP SEQUENCE OF 4993-5127 FROM N.A.
 RC STRAIN=Isolate Vietnam;
 RA Emery S., Erdman D., Peret T., Ksiazek T.;
 RL Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
 RN [6]
 RP SEQUENCE OF 4993-5136 FROM N.A.
 RC STRAIN=Isolate Taiwan;
 RA Lin J.-H., Chiu S.-C., Yang J.-Y., Wang S.-F., Chen H.-Y.;
 RT "Detection of a novel human coronavirus in a severe acute respiratory
 RT syndrome patient in Taiwan.";
 RL Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
 CC -!- FUNCTION: The replicase polyprotein of coronaviruses is a
 CC multifunctional protein: it contains the activities necessary for
 CC the transcription of negative stranded RNA, leader RNA, subgenomic
 CC mRNAs and progeny virion RNA as well as proteinases responsible
 CC for the cleavage of the polyprotein into functional products (By
 CC similarity).
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA}(N).
 CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins
 CC (By similarity).
 CC -!- MISCELLANEOUS: This protein is translated as a 1A-1B polyprotein
 CC by a ribosomal frameshifting mechanism (By similarity).
 CC -!- SIMILARITY: Contains 1 peptidase family C16 domain.
 CC -!- SIMILARITY: Contains 1 peptidase family C30 domain.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AY278741; AAP13442.1; -.
 DR EMBL; AY278741; AAP13439.1; -.
 DR EMBL; AY278741; AAP13440.1; ALT_SEQ.
 DR EMBL; AY274119; -; NOT_ANNOTATED_CDS.
 DR EMBL; AY278554; AAP13566.1; -.
 DR EMBL; AY278554; AAP13575.1; -.
 DR EMBL; AY278491; -; NOT_ANNOTATED_CDS.
 DR EMBL; AY269391; AAP04003.1; -.
 DR EMBL; AY268049; AAP04587.1; -.
 DR InterPro; IPR002589; Alpp.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSvir.
 DR InterPro; IPR002877; FtsJ.
 DR Pfam; PF01661; Alpp; 1.
 DR Pfam; PF01728; FtsJ; 1.
 DR SMART; SM00506; Alpp; 1.
 KW Polyprotein; Transferase; RNA-directed RNA polymerase; Thiol protease;
 KW Hydrolase; Helicase; ATP-binding.
 FT DOMAIN 1 179 LEADER PROTEIN (POTENTIAL).
 FT DOMAIN 180 818 P65 HOMOLOG (POTENTIAL).
 FT DOMAIN ? ? PAPAINE-LIKE PROTEINASE (POTENTIAL).
 FT DOMAIN 3240 3547 3C-LIKE PROTEINASE (POTENTIAL).
 FT DOMAIN 3548 3836 HD2/NSP3 (POTENTIAL).
 FT DOMAIN 3837 3919 NSP4 (POTENTIAL).
 FT DOMAIN 3920 4117 NSP5 (POTENTIAL).
 FT DOMAIN 4118 4229 NSP6 (POTENTIAL).
 FT DOMAIN 4230 4369 GROWTH FACTOR-LIKE (POTENTIAL).
 FT DOMAIN 4370 5301 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT DOMAIN 5302 5902 HELICASE (POTENTIAL).
 FT DOMAIN 5903 6429 NSP11 (POTENTIAL).
 FT DOMAIN 6430 6775 NSP12 (POTENTIAL).
 FT DOMAIN 6776 7073 NSP13 (POTENTIAL).
 FT ACT_SITE 1909 1909 POTENTIAL.
 FT NP_BIND 5583 5590 ATP (POTENTIAL).
 FT DOMAIN 930 933 POLY-GLU.
 FT DOMAIN 937 942 POLY-GLU.
 FT DOMAIN 974 979 POLY-GLU.
 FT DOMAIN 2210 2213 POLY-LEU.
 FT DOMAIN 3766 3769 POLY-CYS.
 FT VARIANT 2552 2552 V -> A (in isolates Tor2, CUHK-W1 and
 FT HKU-39849).
 FT VARIANT 2556 2556 D -> N (in isolate HKU-39849).
 FT VARIANT 2708 2708 S -> T (in isolate HKU-39849).
 FT VARIANT 2718 2718 R -> T (in isolate HKU-39849).
 FT VARIANT 3047 3047 V -> A (in isolate CUHK-W1).
 FT VARIANT 3072 3072 V -> A (in isolate CUHK-W1).
 FT VARIANT 4379 4382 RVCG -> GFAV (in ORF1A).
 FT VARIANT 5131 5131 A -> G (in isolate Taiwan).
 FT VARIANT 5134 5135 CY -> VL (in isolate Taiwan).

FT VARIANT 5767 5767 D -> E (in isolate CUHK-W1).
 FT VARIANT 6778 6778 Q -> R (in isolate Tor2).
 FT VARIANT 6883 6883 D -> Y (in isolate Tor2).
 SQ SEQUENCE 7073 AA; 790270 MW; A91B3CE920E69D4C CRC64;

Query Match 66.7%; Score 36; DB 1; Length 7073;
 Best Local Similarity 66.7%; Pred. No. 3e+02;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
 |||: |||
 Db 5309 CNSQTSRLC 5317

RESULT 7

NHAR_VIBCH

ID NHAR_VIBCH STANDARD; PRT; 296 AA.
 AC P52692; Q9JMP8;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcriptional activator protein nhaR (Na⁺/H⁺ antiporter regulatory protein).
 GN NHAR OR HLYT OR VC0677.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Classical Inaba Z17561 / Serotype O1;
 RX MEDLINE=94049116; PubMed=8231807;
 RA Williams S.G., Attridge S.R., Manning P.A.;
 RT "The transcriptional activator HlyU of Vibrio cholerae: nucleotide sequence and role in virulence gene expression.";
 RL Mol. Microbiol. 9:751-760(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=El Tor O17 / Serotype O1;
 RX MEDLINE=98117066; PubMed=9457888;
 RA Williams S.G., Carmel-Harel O., Manning P.A.;
 RT "A functional homolog of Escherichia coli NhaR in Vibrio cholerae.";
 RL J. Bacteriol. 180:762-765(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=El Tor N16961 / Serotype O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.";
 RL Nature 406:477-483(2000).

CC -!- FUNCTION: PLAYS A ROLE IN THE POSITIVE REGULATION OF NHAA
CC (PROBABLE).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; X66866; CAA47335.1; -.
DR EMBL; AJ002395; CAA05371.1; -.
DR EMBL; AE004154; AAF93842.1; -.
DR PIR; S37312; S37312.
DR TIGR; VC0677; -.
DR InterPro; IPR000847; HTH_LysR.
DR Pfam; PF00126; HTH_1; 1.
DR PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
KW Transcription regulation; DNA-binding; Activator; Complete proteome.
FT DNA_BIND 21 40 H-T-H MOTIF (POTENTIAL).
SQ SEQUENCE 296 AA; 33559 MW; C7830B4B532DBC0C CRC64;

Query Match 64.8%; Score 35; DB 1; Length 296;
Best Local Similarity 62.5%; Pred. No. 15;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 NSRLHLRC 9
 ::|:||||
Db 119 DNRIHLRC 126

RESULT 8

PHNM_ECOLI

ID PHNM_ECOLI STANDARD; PRT; 378 AA.
AC P16689;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE PhnM protein.
GN PHNM OR B4095.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=91193228; PubMed=1840580;
RA Makino K., Kim S.K., Shinagawa H., Amemura M., Nakata A.;
RT "Molecular analysis of the cryptic and functional phn operons for
RT phosphonate use in Escherichia coli K-12."
RL J. Bacteriol. 173:2665-2672(1991).
RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=95334362; PubMed=7610040;
 RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
 RA Blattner F.R.;
 RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
 RT region from 92.8 through 100 minutes.";
 RL Nucleic Acids Res. 23:2105-2119(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B;
 RX MEDLINE=90170953; PubMed=2155230;
 RA Chen C.-M., Ye Q.-Z., Zhu Z., Wanner B.L., Walsh C.T.;
 RT "Molecular biology of carbon-phosphorus bond cleavage. Cloning and
 RT sequencing of the phn (psiD) genes involved in alkylphosphonate
 RT uptake and C-P lyase activity in Escherichia coli B.";
 RL J. Biol. Chem. 265:4461-4471(1990).
 CC -!- FUNCTION: BELONGS TO AN OPERON INVOLVED IN ALKYLPHOSPHONATE
 CC UPTAKE AND C-P LYASE. EXACT FUNCTION NOT KNOWN.
 CC -!- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF STRAIN K12.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D90227; BAA14273.1; -.
 DR EMBL; U14003; AAA96994.1; -.
 DR EMBL; AE000482; AAC77056.1; -.
 DR EMBL; J05260; AAA24352.1; -.
 DR PIR; S56323; S56323.
 DR EcoGene; EG10722; phnM.
 DR InterPro; IPR006680; Amidohydro_1.
 DR Pfam; PF01979; Amidohydro_1; 1.
 DR ProDom; PD000518; Urease; 1.
 KW Alkylphosphonate uptake; Complete proteome.
 FT VARIANT 318 318 Q -> E (IN STRAIN B).
 SQ SEQUENCE 378 AA; 42010 MW; 28CC9C5C77EAD37D CRC64;

Query Match 64.8%; Score 35; DB 1; Length 378;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 RLHLRC 9
 |||||
 Db 132 RLHLRC 137

RESULT 9
 TRYS_CRIFA
 ID TRYS_CRIFA STANDARD; PRT; 652 AA.
 AC O60993;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Trypanothione synthetase (EC 6.3.1.9) (Cf-TS).
 GN TRS.
 OS Crithidia fasciculata.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Crithidia.
 OX NCBI_TaxID=5656;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 168-191; 232-255; 437-450;
 RP 457-474; 592-610 AND 617-630.
 RC STRAIN=HS6;
 RX MEDLINE=98344022; PubMed=9677355;
 RA Tetaud E., Manai F., Barrett M.P., Nadeau K., Walsh C.T.,
 RA Fairlamb A.H.;
 RT "Cloning and characterization of the two enzymes responsible for
 RT trypanothione biosynthesis in Crithidia fasciculata.";
 RL J. Biol. Chem. 273:19383-19390(1998).
 RN [2]
 RP SEQUENCE OF 168-191; 232-255; 437-450; 457-474; 592-610 AND 617-630,
 RP AND CHARACTERIZATION.
 RX MEDLINE=93278303; PubMed=1304372;
 RA Smith K., Nadeau K., Bradley M., Walsh C., Fairlamb A.H.;
 RT "Purification of glutathionylspermidine and trypanothione synthetases
 RT from Crithidia fasciculata.";
 RL Protein Sci. 1:874-883(1992).
 RN [3]
 RP SEQUENCE OF 17-36; 224-229; 235-242; 515-533; 550-561 AND 617-629, AND
 RP CHARACTERIZATION.
 RX MEDLINE=97277330; PubMed=9115252;
 RA Koenig K., Menge U., Kiess M., Wray V., Flohe L.;
 RT "Convenient isolation and kinetic mechanism of glutathionylspermidine
 RT synthetase from Crithidia fasciculata.";
 RL J. Biol. Chem. 272:11908-11915(1997).
 CC -!- FUNCTION: Conjugates glutathione (gamma-Glu-Cys-Gly) and
 CC glutathionylspermidine to form trypanothione (N(1),N(8)-
 CC bis(glutathionyl)spermidine), which is involved in maintaining
 CC intracellular thiol redox and in defense against oxidants.
 CC -!- CATALYTIC ACTIVITY: Gamma-L-glutamyl-L-cysteinyl-glycine + N(1)-
 CC (gamma-L-glutamyl-L-cysteinyl-glycyl)-spermidine + ATP =
 CC N(1),N(8)-bis-(gamma-L-glutamyl-L-cysteinyl-glycyl)-spermidine +
 CC ADP + phosphate.
 CC -!- COFACTOR: Magnesium.
 CC -!- PTM: The N-Terminal is blocked.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF006615; AAC39132.1; -.
 DR InterPro; IPR005494; GSP_synth.
 DR Pfam; PF05257; AXE; 1.
 DR Pfam; PF03738; GSP_synth; 1.
 KW Ligase; Magnesium.
 SQ SEQUENCE 652 AA; 74516 MW; 321BE90D39EEEA80 CRC64;

Query Match 64.8%; Score 35; DB 1; Length 652;
Best Local Similarity 55.6%; Pred. No. 36;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CNSRLHLRC 9
|: |||
Db 268 CDHEFHLRC 276

RESULT 10

UBR1_KLULA

ID UBR1_KLULA STANDARD; PRT; 1941 AA.
AC O60014;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE N-end-recognizing protein (Ubiquitin-protein ligase E3 component) (N-
DE recognin).
GN UBR1.
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RA Waller P.R.H., Varshavsky A.;
RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: RECOGNITION COMPONENT OF THE N-END RULE PATHWAY. BINDS
CC TO PROTEINS BEARING AMINO-TERMINAL RESIDUES THAT ARE DESTABILIZING
CC ACCORDING TO THE N-END RULE, BUT DOES NOT BIND TO OTHERWISE
CC IDENTICAL PROTEINS BEARING STABILIZING AMINO-TERMINAL RESIDUES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF061554; AAC15841.1; -.
DR PIR; T30554; T30554.
DR InterPro; IPR003126; Znf_Nrecognin.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF02207; zf-UBR1; 1.
DR SMART; SM00184; RING; 1.
DR SMART; SM00396; ZnF_UBR1; 1.
KW Ligase; Ubl conjugation pathway.
SQ SEQUENCE 1941 AA; 223682 MW; 37C2E1BCA0803268 CRC64;

Query Match 64.8%; Score 35; DB 1; Length 1941;
Best Local Similarity 55.6%; Pred. No. 1.1e+02;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CNSRLHLRC 9
|| :| ||

Db 1271 CNHAVHYRC 1279

RESULT 11

RL37_METTH

ID RL37_METTH STANDARD; PRT; 60 AA.
AC O26744;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L37e.
GN RPL37E OR MTH648.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics."
RL J. Bacteriol. 179:7135-7155(1997).
CC -!- SIMILARITY: BELONGS TO THE L37E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000845; AAB85153.1; -.
DR PIR; B69186; B69186.
DR HAMAP; MF_00547; -; 1.
DR InterPro; IPR001569; Ribosomal_L37E.
DR Pfam; PF01907; Ribosomal_L37e; 1.
DR ProDom; PD005132; Ribosomal_L37E; 1.
DR PROSITE; PS01077; RIBOSOMAL_L37E; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 60 AA; 7123 MW; 3B7026A579EAC9D5 CRC64;

Query Match 63.0%; Score 34; DB 1; Length 60;
Best Local Similarity 62.5%; Pred. No. 4.2;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 NSRLHLRC 9
| ||:|
Db 11 NKNLHIRC 18

RESULT 12

ECR1_AERPE

ID ECR1_AERPE STANDARD; PRT; 235 AA.
AC Q9YC02;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable exosome complex RNA-binding protein 1.
GN APE1448.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococcaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
CC -!- FUNCTION: Probably involved in degradation of a variety of RNA
CC species; could act a RNA-binding component of the exosome
CC (Potential).
CC -!- SUBUNIT: Component of the archaeal exosome multienzyme
CC ribonuclease complex (Potential).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: Contains 1 KH domain.
CC -!- SIMILARITY: Contains 1 S1 motif domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP000061; BAA80446.1; -.
DR PIR; H72623; H72623.
DR HAMAP; MF_00623; -; 1.
DR InterPro; IPR004087; KH_dom.
DR InterPro; IPR003029; S1.
DR Pfam; PF00013; KH; 1.
DR SMART; SM00322; KH; 1.
DR SMART; SM00316; S1; 1.
DR PROSITE; PS50084; KH_TYPE_1; FALSE_NEG.
DR PROSITE; PS50126; S1; 1.
KW Exosome; RNA-binding; Complete proteome.
FT DOMAIN 67 139 S1 MOTIF.

FT DOMAIN 147 206 KH.
SQ SEQUENCE 235 AA; 26060 MW; 70A79A5EB0BF8CE7 CRC64;

Query Match 63.0%; Score 34; DB 1; Length 235;
Best Local Similarity 62.5%; Pred. No. 18;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 NSRLHLRC 9
| | : | | |
Db 186 NGRIHLEC 193

RESULT 13

YQHQ_BACSU

ID YQHQ_BACSU STANDARD; PRT; 318 AA.
AC P54515;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yqhQ.
GN YQHQ.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE=97124195; PubMed=8969508;
RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,
RA Kobayashi Y.;
RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of
RT the Bacillus subtilis genome containing the skin element and many
RT sporulation genes.";
RL Microbiology 142:3103-3111(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,

RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*.";
 RL Nature 390:249-256(1997).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D84432; BAA12554.1; -.
 DR EMBL; Z99116; CAB14380.1; -.
 DR PIR; H69959; H69959.
 DR SubtiList; BG11705; yqhQ.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 112 132 POTENTIAL.
 FT TRANSMEM 147 167 POTENTIAL.
 FT TRANSMEM 209 229 POTENTIAL.
 FT TRANSMEM 237 257 POTENTIAL.
 SQ SEQUENCE 318 AA; 36001 MW; 15619B40274BB716 CRC64;

Query Match 63.0%; Score 34; DB 1; Length 318;
 Best Local Similarity 85.7%; Pred. No. 25;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 SRLHLRC 9
 |||| ||
 Db 203 SRLHYRC 209

RESULT 14

AS14_MOUSE
 ID AS14_MOUSE STANDARD; PRT; 433 AA.
 AC Q8VHS7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Ankyrin repeat and SOCS box containing protein 14 (ASB-14).
 GN ASB14.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Kile B.T., Nicola N.A.;
 RT "SOCS box proteins.";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
 CC -!- SIMILARITY: Contains 9 ANK repeats.
 CC -!- SIMILARITY: Contains 1 SOCS box domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF403042; AAL57361.1; -.
 DR MGD; MGI:2655107; Asb14.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR001496; SOCS.
 DR Pfam; PF00023; ank; 8.
 DR SMART; SM00248; ANK; 8.
 DR PROSITE; PS50088; ANK_REPEAT; 6.
 DR PROSITE; PS50297; ANK_REP_REGION; 1.
 DR PROSITE; PS50225; SOCS; 1.
 KW ANK repeat; Repeat.
 FT REPEAT 1 14 ANK 1.
 FT REPEAT 18 47 ANK 2.
 FT REPEAT 51 80 ANK 3.
 FT REPEAT 94 123 ANK 4.
 FT REPEAT 127 156 ANK 5.
 FT REPEAT 159 188 ANK 6.
 FT REPEAT 201 230 ANK 7.
 FT REPEAT 231 260 ANK 8.
 FT REPEAT 262 295 ANK 9.
 FT DOMAIN 367 422 SOCS BOX.
 SQ SEQUENCE 433 AA; 48317 MW; 6BCAD1AC2B2BB080 CRC64;

Query Match 63.0%; Score 34; DB 1; Length 433;
 Best Local Similarity 66.7%; Pred. No. 35;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CNSRLHLRC 9
 | ||||
 Db 391 CMGRLLRLRC 399

RESULT 15

NO70_SOYBN

ID NO70_SOYBN STANDARD; PRT; 485 AA.
 AC Q02920;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Early nodulin 70.
 OS Glycine max (Soybean).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

OX NCBI_TaxID=3847;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Akisengoku;
 RX MEDLINE=93241143; PubMed=7683079;
 RA Kouchi H., Hata S.;
 RT "Isolation and characterization of novel nodulin cDNAs representing
 RT genes expressed at early stages of soybean nodule development.";
 RL Mol. Gen. Genet. 238:106-119(1993).
 RN [2]
 RP SIMILARITY TO SULFATE PERMEASES.
 RX MEDLINE=94188926; PubMed=8140616;
 RA Sandal N.N., Marcker K.A.;
 RT "Similarities between a soybean nodulin, Neurospora crassa sulphate
 RT permease II and a putative human tumour suppressor.";
 RL Trends Biochem. Sci. 19:19-19(1994).
 CC -!- FUNCTION: POSSIBLE SULFATE TRANSPORTER.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED AT EARLY STAGES OF NODULE
 CC DEVELOPMENT.
 CC -!- SIMILARITY: BELONGS TO THE SLC26A FAMILY OF TRANSPORTERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D13505; BAA02723.1; -.
 DR PIR; S34800; S34800.
 DR InterPro; IPR001902; Sulph_transpt.
 DR Pfam; PF00916; Sulfate_transp; 1.
 DR TIGRFAMs; TIGR00815; sulP; 1.
 DR PROSITE; PS01130; SLC26A; 1.
 KW Nodulation; Transmembrane; Transport.
 FT TRANSMEM 74 94 POTENTIAL.
 FT TRANSMEM 97 117 POTENTIAL.
 FT TRANSMEM 121 141 POTENTIAL.
 FT TRANSMEM 152 172 POTENTIAL.
 FT TRANSMEM 186 206 POTENTIAL.
 FT TRANSMEM 232 252 POTENTIAL.
 FT TRANSMEM 262 282 POTENTIAL.
 FT TRANSMEM 321 341 POTENTIAL.
 FT TRANSMEM 361 381 POTENTIAL.
 FT TRANSMEM 400 420 POTENTIAL.
 FT TRANSMEM 421 441 POTENTIAL.
 FT TRANSMEM 455 475 POTENTIAL.
 SQ SEQUENCE 485 AA; 52945 MW; 3738B6F64383BCB5 CRC64;

Query Match 63.0%; Score 34; DB 1; Length 485;
 Best Local Similarity 75.0%; Pred. No. 40;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLHLR 8
 | || |||

Db 477 CRSRYHLR 484

Search completed: November 13, 2003, 09:46:29
Job time : 6.15625 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 23.7188 Seconds
(without alignments)
97.917 Million cell updates/sec

Title: US-09-228-866-1
Perfect score: 54
Sequence: 1 CNSRLHLRC 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query	Match Length			
1	44	81.5	142	10	Q94DJ7	Q94dj7 oryza sativ
2	40	74.1	1327	4	O60859	O60859 homo sapien
3	40	74.1	1327	11	Q9R114	Q9r114 mus musculu
4	40	74.1	1332	4	Q8IY17	Q8iy17 homo sapien
5	39	72.2	1348	5	Q8I2K9	Q8i2k9 plasmodium
6	38	70.4	220	10	Q9LS99	Q9ls99 arabidopsis
7	38	70.4	309	10	Q8S3N1	Q8s3n1 arabidopsis
8	38	70.4	315	10	Q9XGC2	Q9xgc2 vitis vinif
9	38	70.4	451	2	Q9LAB9	Q9lab9 pseudomonas
10	38	70.4	514	16	Q9PBA3	Q9pba3 xylella fas
11	38	70.4	669	5	Q8ISE4	Q8ise4 drosophila
12	38	70.4	669	5	Q8ISD4	Q8isd4 drosophila
13	38	70.4	675	5	Q8ISE3	Q8ise3 drosophila
14	38	70.4	675	5	Q8ISE2	Q8ise2 drosophila
15	38	70.4	675	5	Q8ISE1	Q8isel drosophila
16	38	70.4	675	5	Q8I0D9	Q8i0d9 drosophila
17	38	70.4	675	5	Q8I086	Q8i086 drosophila
18	38	70.4	739	5	Q9W4Z3	Q9w4z3 drosophila
19	38	70.4	782	5	Q9NEH9	Q9neh9 drosophila
20	37	68.5	464	10	Q9SUA5	Q9sua5 arabidopsis
21	37	68.5	574	10	Q9M8Z9	Q9m8z9 arabidopsis
22	37	68.5	736	6	Q28482	Q28482 macaca fasc
23	37	68.5	1430	5	Q23541	Q23541 caenorhabdi
24	36	66.7	146	10	O64874	O64874 arabidopsis
25	36	66.7	199	4	Q8NHT9	Q8nht9 homo sapien
26	36	66.7	259	3	Q9P3T8	Q9p3t8 schizosacch
27	36	66.7	332	13	Q98U07	Q98u07 pseudotylos
28	36	66.7	332	13	Q98U08	Q98u08 platybelone
29	36	66.7	333	13	Q9DF04	Q9df04 strongylura
30	36	66.7	333	13	Q9DF15	Q9df15 platybelone
31	36	66.7	333	13	Q9DF08	Q9df08 strongylura
32	36	66.7	333	13	Q9DF10	Q9df10 potamorrhap
33	36	66.7	333	13	Q9DF14	Q9df14 potamorrhap
34	36	66.7	333	13	Q9DF01	Q9df01 belonion ap
35	36	66.7	333	13	Q9DD82	Q9dd82 potamorrhap
36	36	66.7	333	13	Q9DD51	Q9dd51 pseudotylos
37	36	66.7	333	13	Q9DD50	Q9dd50 belonion di
38	36	66.7	333	13	Q9DF03	Q9df03 strongylura
39	36	66.7	333	13	Q9DF16	Q9df16 strongylura
40	36	66.7	333	13	Q9DF12	Q9df12 strongylura
41	36	66.7	333	13	Q9DD64	Q9dd64 strongylura
42	36	66.7	333	13	Q9DD35	Q9dd35 strongylura
43	36	66.7	333	13	Q9DF13	Q9df13 potamorrhap
44	36	66.7	333	13	Q9DF05	Q9df05 strongylura
45	36	66.7	333	13	Q9DF02	Q9df02 strongylura

ALIGNMENTS

RESULT 1
Q94DJ7

ID Q94DJ7 PRELIMINARY; PRT; 142 AA.
AC Q94DJ7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE P0514H03.16 protein.
GN P0514H03.16.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0514H03.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AP003275; BAB63652.1; -.
DR Gramene; Q94DJ7; -.
SQ SEQUENCE 142 AA; 14769 MW; 507041848FEEA897CRC64;

Query Match 81.5%; Score 44; DB 10; Length 142;
Best Local Similarity 77.8%; Pred. No. 0.32;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
| | | | |
Db 50 CQGRHLHLRC 58

RESULT 2

O60859

ID O60859 PRELIMINARY; PRT; 1327 AA.
AC O60859;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Neuropathy target esterase.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98244804; PubMed=9576844;
RA Lush M.J., Li Y., Read D.J., Willis A.C., Glynn P.;
RT "Neuropathy Target Esterase (NTE) and a homologous Drosophila
RT neurodegeneration-associated mutant protein contain a novel domain
RT conserved from bacteria to man.";
RL Biochem. J. 332:1-4(1998).
DR EMBL; AJ004832; CAA06164.1; -.
DR InterPro; IPR000595; cNMP_binding.
DR InterPro; IPR002641; Patatin.
DR InterPro; IPR001423; UPF0028.

DR Pfam; PF00027; cNMP_binding; 3.
 DR Pfam; PF01734; Patatin; 1.
 DR SMART; SM00100; cNMP; 2.
 DR PROSITE; PS50042; CNMP_BINDING_3; 3.
 DR PROSITE; PS01237; UPF0028; 1.
 SQ SEQUENCE 1327 AA; 146215 MW; E823248C9B29DD84 CRC64;

Query Match 74.1%; Score 40; DB 4; Length 1327;
 Best Local Similarity 66.7%; Pred. No. 16;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
 |: |||||
 Db 880 CSGHLHLRC 888

RESULT 3

Q9R114

ID Q9R114 PRELIMINARY; PRT; 1327 AA.
 AC Q9R114;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Neuropathy target esterase homolog.
 GN NTE.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Balb/c;
 RA Kretzschmar D., Stempf1 T., Moser M.;
 RT "Cloning of murine sws/NTE."
 RL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AF173829; AAD51700.1; -.
 DR MGD; MGI:1354723; Nte.
 DR InterPro; IPR000595; cNMP_binding.
 DR InterPro; IPR002641; Patatin.
 DR InterPro; IPR001423; UPF0028.
 DR Pfam; PF00027; cNMP_binding; 3.
 DR Pfam; PF01734; Patatin; 1.
 DR SMART; SM00100; cNMP; 2.
 DR PROSITE; PS50042; CNMP_BINDING_3; 3.
 DR PROSITE; PS01237; UPF0028; 1.
 SQ SEQUENCE 1327 AA; 146561 MW; 1824D4F6BA6BEC70 CRC64;

Query Match 74.1%; Score 40; DB 11; Length 1327;
 Best Local Similarity 66.7%; Pred. No. 16;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
 |: |||||
 Db 880 CSGHLHLRC 888

RESULT 4

Q8IY17

ID Q8IY17 PRELIMINARY; PRT; 1332 AA.
AC Q8IY17;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Neuropathy target esterase (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC038229; AAH38229.1; -.
FT NON_TER 1 1
SQ SEQUENCE 1332 AA; 146818 MW; 7B1C72CAB920B97A CRC64;

Query Match 74.1%; Score 40; DB 4; Length 1332;
Best Local Similarity 66.7%; Pred. No. 16;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
|: |||||
Db 885 CSGHLHLRC 893

RESULT 5

Q8I2K9

ID Q8I2K9 PRELIMINARY; PRT; 1348 AA.
AC Q8I2K9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Diacylglycerol kinase, putative (EC 2.7.1.107).
GN PFI1485C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabbinoiwitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,

RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
 RA Sulston J.E., Craig A., Newbold C., Barrell B.G;
 RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."
 RL Nature 419:527-531(2002).
 DR EMBL; AL929358; CAD51983.1; -.
 KW Kinase; Transferase.
 SQ SEQUENCE 1348 AA; 158972 MW; 7523D6F052DB18FD CRC64;

Query Match 72.2%; Score 39; DB 5; Length 1348;
 Best Local Similarity 66.7%; Pred. No. 25;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
 || ||||
 Db 222 CNKYFHLRC 230

RESULT 6

Q9LS99

ID Q9LS99 PRELIMINARY; PRT; 220 AA.

AC Q9LS99;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Contains similarity to RING zinc finger protein.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Columbia;

RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;

RL Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Columbia;

RX MEDLINE=20277480; PubMed=10819329;

RA Nakamura Y.;

RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence

RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC

RT clones.";

RL DNA Res. 7:131-135(2000).

RN [3]

RP SEQUENCE FROM N.A.

RA Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,

RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;

RT "Full-length messenger RNA sequences greatly improve genome

RT annotation.";

RL Genome Biol. 0:0-0(2002).

RN [4]

RP SEQUENCE FROM N.A.

RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,

RA Feldmann K.;

RT "Full-Length cDNA from Arabidopsis thaliana.";

RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.

CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
 DR EMBL; AB026654; BAB01804.1; -.
 DR EMBL; AY086917; AAM64481.1; -.
 DR HSSP; P28990; 1CHC.
 DR InterPro; IPR001841; Znf_ring.
 DR Pfam; PF00097; zf-C3HC4; 1.
 DR PROSITE; PS50089; ZF_RING_2; 1.
 KW Metal-binding; Zinc; Zinc-finger.
 SQ SEQUENCE 220 AA; 24463 MW; F63F08AEACA4494D CRC64;

Query Match 70.4%; Score 38; DB 10; Length 220;
 Best Local Similarity 66.7%; Pred. No. 7.7;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
 || ||||
 Db 149 CNHGFHLRC 157

RESULT 7

Q8S3N1

ID Q8S3N1 PRELIMINARY; PRT; 309 AA.
 AC Q8S3N1;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Ring finger E3 ligase SINAT5.
 GN SINAT5.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Xie Q., Chua N.H.;
 RT "SINAT 5, a RING E3 ubiquitin protein ligase, promotes post-
 RT translational degradation of NAC 1 to attenuate auxin signals."
 RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AF480944; AAM11573.1; -.
 DR InterPro; IPR004162; Sina.
 DR InterPro; IPR001841; Znf_ring.
 DR Pfam; PF03145; Sina; 1.
 DR PROSITE; PS50089; ZF_RING_2; 1.
 KW Ligase.
 SQ SEQUENCE 309 AA; 35008 MW; 3908E2353BB57AAF CRC64;

Query Match 70.4%; Score 38; DB 10; Length 309;
 Best Local Similarity 66.7%; Pred. No. 11;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
 | ||:| ||
 Db 70 CKSRVHNRC 78

RESULT 8

Q9XGC2

ID Q9XGC2 PRELIMINARY; PRT; 315 AA.
 AC Q9XGC2;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE SINA1p.
 OS Vitis vinifera (Grape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae;
 OC Vitis.
 OX NCBI_TaxID=29760;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Optima;
 RA Brehm I., Korfei M., Preisig-Mueller R., Kindl H.;
 RT "A nuclear localized zinc finger protein found in a plant is
 RT homologous to the Drosophila signal transducing factor seven in
 RT absentia.";
 RL Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; Y18471; CAB40577.1; -.
 DR InterPro; IPR004162; Sina.
 DR InterPro; IPR001841; Znf_ring.
 DR Pfam; PF03145; Sina; 1.
 DR PROSITE; PS50089; ZF_RING_2; 1.
 SQ SEQUENCE 315 AA; 35838 MW; BC49A24384F6D028 CRC64;

Query Match 70.4%; Score 38; DB 10; Length 315;
 Best Local Similarity 66.7%; Pred. No. 11;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
 | ||:| ||
 Db 78 CKSRVHNRC 86

RESULT 9

Q9LAB9

ID Q9LAB9 PRELIMINARY; PRT; 451 AA.
 AC Q9LAB9;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Polyurethanase lipase A.
 GN PULA.
 OS Pseudomonas fluorescens.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=294;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ruiz C., Vega R., Howard G.T.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AF144089; AAF66684.1; -.
 DR InterPro; IPR001343; Hemlysn_Ca_bind.
 DR InterPro; IPR000734; Lipase.
 DR Pfam; PF00353; hemolysinCabind; 3.

DR PRINTS; PR00313; CABNDNGRPT.
 DR PROSITE; PS00330; HEMOLYSIN_CALCIIUM; 1.
 DR PROSITE; PS00120; LIPASE_SER; 1.
 SQ SEQUENCE 451 AA; 48187 MW; 1164AAE73BFD0CA3 CRC64;

Query Match 70.4%; Score 38; DB 2; Length 451;
 Best Local Similarity 66.7%; Pred. No. 15;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CNSRLHLRC 9
 | : ||| ||
 Db 128 CDHRLHRRRC 136

RESULT 10

Q9PBA3

ID Q9PBA3 PRELIMINARY; PRT; 514 AA.
 AC Q9PBA3;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Periplasmic protease.
 GN XF2241.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9a5c;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;

RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
 RL Nature 406:151-159(2000).
 CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL; AE004037; AAF85040.1; -.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR InterPro; IPR000126; Ser_proteas_V8.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR PRINTS; PR00839; V8PROTEASE.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS50106; PDZ; 1.
 KW Hydrolase; Serine protease; Complete proteome.
 SQ SEQUENCE 514 AA; 54140 MW; 707C23FD3C82BE4C CRC64;

Query Match 70.4%; Score 38; DB 16; Length 514;
 Best Local Similarity 75.0%; Pred. No. 17;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 NSRLHLRC 9
 |||:| ||
 Db 2 NSRIHTRC 9

RESULT 11

Q8ISE4

ID Q8ISE4 PRELIMINARY; PRT; 669 AA.
 AC Q8ISE4;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Polehole (Fragment).
 GN PH.
 OS Drosophila mauritiana (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7226;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAU11;
 RA Riley R.M., Jin W., Gibson G.;
 RT "Contrasting selection pressures on components of the Ras-mediated
 RT signal transduction pathway in Drosophila."
 RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AY135030; AAN17540.1; -.
 FT NON_TER 1 1
 SQ SEQUENCE 669 AA; 75788 MW; 7A3E8729F9425927 CRC64;

Query Match 70.4%; Score 38; DB 5; Length 669;
 Best Local Similarity 66.7%; Pred. No. 21;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
 || || || ||

Db 182 CNFRFHQRC 190

RESULT 12

Q8ISD4

ID Q8ISD4 PRELIMINARY; PRT; 669 AA.
AC Q8ISD4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Polehole (Fragment).
GN PH.
OS Drosophila simulans (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SIM31;
RA Riley R.M., Jin W., Gibson G.;
RT "Contrasting selection pressures on components of the Ras-mediated
RT signal transduction pathway in Drosophila."
RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY135135; AAN17563.1; -.
FT NON_TER 1 1
SQ SEQUENCE 669 AA; 75772 MW; C6204C078263B03C CRC64;

Query Match 70.4%; Score 38; DB 5; Length 669;
Best Local Similarity 66.7%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
|||
Db 182 CNFRFHQRC 190

RESULT 13

Q8ISE3

ID Q8ISE3 PRELIMINARY; PRT; 675 AA.
AC Q8ISE3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Polehole (Fragment).
GN PH.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RA Riley R.M., Jin W., Gibson G.;
RT "Contrasting selection pressures on components of the Ras-mediated
RT signal transduction pathway in Drosophila.";

RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY135031; AAN17541.1; -.
FT NON_TER 1 1
SQ SEQUENCE 675 AA; 76488 MW; 29449E17C54A6125 CRC64;

Query Match 70.4%; Score 38; DB 5; Length 675;
Best Local Similarity 66.7%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
|||
Db 188 CNFRFHQRC 196

RESULT 14

Q8ISE2

ID Q8ISE2 PRELIMINARY; PRT; 675 AA.
AC Q8ISE2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Polehole (Fragment).
GN PH.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Reids2;
RA Riley R.M., Jin W., Gibson G.;
RT "Contrasting selection pressures on components of the Ras-mediated
RT signal transduction pathway in Drosophila."
RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY135038; AAN17548.1; -.
FT NON_TER 1 1
SQ SEQUENCE 675 AA; 76541 MW; FE2ACB49901B383E CRC64;

Query Match 70.4%; Score 38; DB 5; Length 675;
Best Local Similarity 66.7%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLHLRC 9
|||
Db 188 CNFRFHQRC 196

RESULT 15

Q8ISE1

ID Q8ISE1 PRELIMINARY; PRT; 675 AA.
AC Q8ISE1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Polehole (Fragment).
GN PH.

```

OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=5-17-88b#5;
RA   Riley R.M., Jin W., Gibson G.;
RT   "Contrasting selection pressures on components of the Ras-mediated
RT   signal transduction pathway in Drosophila.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY135052; AAN17562.1; -.
FT   NON_TER      1      1
SQ   SEQUENCE     675 AA;  76502 MW;  0B2A538ED51B876E CRC64;

Query Match          70.4%;  Score 38;  DB 5;  Length 675;
Best Local Similarity 66.7%;  Pred. No. 21;
Matches      6;  Conservative      0;  Mismatches      3;  Indels      0;  Gaps      0;

Qy          1 CNSRLHLRC 9
           || || || ||
Db          188 CNFRFHQRC 196

```

Search completed: November 13, 2003, 09:50:55
 Job time : 24.7188 secs

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:39:50 ; Search time 10.6875 Seconds
 (without alignments)
 35.630 Million cell updates/sec

Title: US-09-228-866-2
 Perfect score: 67
 Sequence: 1 CENWWGDVC 9

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : Issued_Patents_AA:*
 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		% Match	Length			
1	67	100.0	9	1	US-08-526-710-2	Sequence 2, Appli
2	67	100.0	9	3	US-08-862-855-2	Sequence 2, Appli
3	67	100.0	9	3	US-09-226-985-2	Sequence 2, Appli
4	67	100.0	9	4	US-09-227-906-2	Sequence 2, Appli
5	41	61.2	206	4	US-09-071-035-272	Sequence 272, App
6	41	61.2	207	4	US-09-071-035-270	Sequence 270, App
7	40	59.7	266	4	US-09-252-991A-17646	Sequence 17646, A
8	40	59.7	372	4	US-09-092-315-13	Sequence 13, Appl
9	40	59.7	425	4	US-09-092-315-6	Sequence 6, Appli
10	40	59.7	425	4	US-09-733-524A-6	Sequence 6, Appli
11	40	59.7	454	4	US-09-092-315-8	Sequence 8, Appli

12	40	59.7	454	4	US-09-733-524A-8	Sequence 8, Appli
13	40	59.7	464	4	US-09-092-315-1	Sequence 1, Appli
14	40	59.7	464	4	US-09-733-524A-1	Sequence 1, Appli
15	40	59.7	476	4	US-09-092-315-5	Sequence 5, Appli
16	40	59.7	476	4	US-09-733-524A-5	Sequence 5, Appli
17	40	59.7	478	4	US-09-092-315-7	Sequence 7, Appli
18	40	59.7	478	4	US-09-733-524A-7	Sequence 7, Appli
19	40	59.7	486	4	US-09-092-315-2	Sequence 2, Appli
20	40	59.7	486	4	US-09-733-524A-2	Sequence 2, Appli
21	40	59.7	636	4	US-09-198-452A-489	Sequence 489, App
22	40	59.7	799	3	US-09-180-439-6	Sequence 6, Appli
23	40	59.7	968	3	US-09-180-439-3	Sequence 3, Appli
24	40	59.7	968	3	US-09-180-439-4	Sequence 4, Appli
25	40	59.7	1016	3	US-09-180-439-8	Sequence 8, Appli
26	40	59.7	1112	3	US-09-353-585-2	Sequence 2, Appli
27	40	59.7	1112	3	US-09-353-585-3	Sequence 3, Appli
28	39	58.2	508	3	US-08-472-240A-16	Sequence 16, Appl
29	39	58.2	832	4	US-09-206-551-21	Sequence 21, Appl
30	39	58.2	853	4	US-09-206-551-13	Sequence 13, Appl
31	39	58.2	872	3	US-07-956-483-12	Sequence 12, Appl
32	38	56.7	86	3	US-08-894-173-52	Sequence 52, Appl
33	38	56.7	86	3	US-08-894-173-53	Sequence 53, Appl
34	38	56.7	86	3	US-09-398-193-52	Sequence 52, Appl
35	38	56.7	86	3	US-09-398-193-53	Sequence 53, Appl
36	38	56.7	239	4	US-09-252-991A-28204	Sequence 28204, A
37	38	56.7	291	4	US-09-252-991A-28294	Sequence 28294, A
38	38	56.7	303	4	US-09-420-786A-3	Sequence 3, Appli
39	38	56.7	437	4	US-09-996-243-355	Sequence 355, App
40	38	56.7	534	4	US-09-199-637A-67	Sequence 67, Appl
41	38	56.7	534	4	US-09-252-991A-26566	Sequence 26566, A
42	38	56.7	1248	3	US-08-726-214-16	Sequence 16, Appl
43	37	55.2	121	4	US-09-252-991A-18698	Sequence 18698, A
44	37	55.2	180	3	US-09-187-331-5	Sequence 5, Appli
45	37	55.2	180	4	US-09-470-946-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1

US-08-526-710-2

; Sequence 2, Application US/08526710

; Patent No. 5622699

; GENERAL INFORMATION:

; APPLICANT: Ruoslahti, Erkki

; APPLICANT: Pasqualini, Renata

; TITLE OF INVENTION: Method of Identifying Molecules That

; TITLE OF INVENTION: Home to a Selected Organ In Vivo

; NUMBER OF SEQUENCES: 44

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell and Flores

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego

; STATE: California

; COUNTRY: United States

; ZIP: 92122

; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,710
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-526-710-2

```

```

Query Match          100.0%; Score 67; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      9; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 CENWWGDVC 9
        |||||
Db      1 CENWWGDVC 9

```

RESULT 2

```

US-08-862-855-2
; Sequence 2, Application US/08862855
; Patent No. 6068829
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/862,855
; FILING DATE:

```



```

; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2621
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-862-855-2

```

```

Query Match          100.0%; Score 67; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      9; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 CENWWGDVC 9
        |||||
Db      1 CENWWGDVC 9

```

```

RESULT 3
US-09-226-985-2
; Sequence 2, Application US/09226985
; Patent No. 6296832
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Molecules That Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/226,985
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-226-985-2

```

```

Query Match          100.0%; Score 67; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      9; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 CENWWGDVC 9
        |||||
Db      1 CENWWGDVC 9

```

RESULT 4

US-09-227-906-2

```

; Sequence 2, Application US/09227906
; Patent No. 6306365
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/227,906
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3424
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-227-906-2

```

```

Query Match          100.0%; Score 67; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      9; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

QY      1 CENWWGDVC 9
        |||||
Db      1 CENWWGDVC 9

```

RESULT 5

```

US-09-071-035-272
; Sequence 272, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 272:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 206 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-272

```

```

Query Match          61.2%; Score 41; DB 4; Length 206;
Best Local Similarity 83.3%; Pred. No. 32;
Matches      5; Conservative 1; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      2 ENWWGD 7
        :|||||
Db      164 KNWWGD 169

```

```

RESULT 6
US-09-071-035-270
; Sequence 270, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 270:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 207 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-270

```

```

Query Match          61.2%; Score 41; DB 4; Length 207;
Best Local Similarity 83.3%; Pred. No. 32;
Matches      5; Conservative 1; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      2 ENWWGD 7
      :|||||
Db      165 KNWWGD 170

```

RESULT 7

```

US-09-252-991A-17646
; Sequence 17646, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17646
; LENGTH: 266
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17646

```

```

Query Match          59.7%; Score 40; DB 4; Length 266;
Best Local Similarity 44.4%; Pred. No. 58;
Matches      4; Conservative 1; Mismatches      4; Indels      0; Gaps      0;

```

```

Qy      1 CENWWGDVC 9
      |: || |
Db      55 CQAWWSQAC 63

```

RESULT 8

US-09-092-315-13
 ; Sequence 13, Application US/09092315
 ; Patent No. 6399337
 ; GENERAL INFORMATION:
 ; APPLICANT: Taylor, Diane E.
 ; APPLICANT: Ge, Zhongming
 ; TITLE OF INVENTION: ALPHA-1, 3-FUCOSYLTRANSFERASE
 ; FILE REFERENCE: 07254/049001
 ; CURRENT APPLICATION NUMBER: US/09/092,315
 ; CURRENT FILING DATE: 1998-06-05
 ; EARLIER APPLICATION NUMBER: US 60/048,857
 ; EARLIER FILING DATE: 1997-06-06
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: FastSEQ for Windows Version 3.0
 ; SEQ ID NO 13
 ; LENGTH: 372
 ; TYPE: PRT
 ; ORGANISM: Helicobacter pylori
 US-09-092-315-13

Query Match 59.7%; Score 40; DB 4; Length 372;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 NWWGD 7
 |||||
 Db 32 NWWGD 36

RESULT 9
 US-09-092-315-6
 ; Sequence 6, Application US/09092315
 ; Patent No. 6399337
 ; GENERAL INFORMATION:
 ; APPLICANT: Taylor, Diane E.
 ; APPLICANT: Ge, Zhongming
 ; TITLE OF INVENTION: ALPHA-1, 3-FUCOSYLTRANSFERASE
 ; FILE REFERENCE: 07254/049001
 ; CURRENT APPLICATION NUMBER: US/09/092,315
 ; CURRENT FILING DATE: 1998-06-05
 ; EARLIER APPLICATION NUMBER: US 60/048,857
 ; EARLIER FILING DATE: 1997-06-06
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: FastSEQ for Windows Version 3.0
 ; SEQ ID NO 6
 ; LENGTH: 425
 ; TYPE: PRT
 ; ORGANISM: Helicobacter pylori
 US-09-092-315-6

Query Match 59.7%; Score 40; DB 4; Length 425;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 NWWGD 7
 |||||
 Db 33 NWWGD 37

RESULT 10

US-09-733-524A-6

```
; Sequence 6, Application US/09733524A
; Patent No. 6534298
; GENERAL INFORMATION:
; APPLICANT: Taylor, Diane E.
; APPLICANT: Ge, Zhongming
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING ALPHA-1,3
; TITLE OF INVENTION: FUCOSYLTRANSFERASES AND EXPRESSION SYSTEMS FOR MAKING
AND
; TITLE OF INVENTION: EXPRESSING THEM (amended)
; FILE REFERENCE: 07254-049002
; CURRENT APPLICATION NUMBER: US/09/733,524A
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 09/092,315
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: US 60/048,857
; PRIOR FILING DATE: 1997-06-06
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 425
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-733-524A-6
```

```
Query Match          59.7%; Score 40; DB 4; Length 425;
Best Local Similarity 100.0%; Pred. No. 93;
Matches    5; Conservative    0; Mismatches    0; Indels    0; Gaps    0;
```

```
QY          3 NWWGD 7
             |||||
Db          33 NWWGD 37
```

RESULT 11

US-09-092-315-8

```
; Sequence 8, Application US/09092315
; Patent No. 6399337
; GENERAL INFORMATION:
; APPLICANT: Taylor, Diane E.
; APPLICANT: Ge, Zhongming
; TITLE OF INVENTION: ALPHA-1, 3-FUCOSYLTRANSFERASE
; FILE REFERENCE: 07254/049001
; CURRENT APPLICATION NUMBER: US/09/092,315
; CURRENT FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: US 60/048,857
; EARLIER FILING DATE: 1997-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-092-315-8
```

Query Match 59.7%; Score 40; DB 4; Length 454;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 NWWGD 7
|||||
Db 32 NWWGD 36

RESULT 12

US-09-733-524A-8
; Sequence 8, Application US/09733524A
; Patent No. 6534298
; GENERAL INFORMATION:
; APPLICANT: Taylor, Diane E.
; APPLICANT: Ge, Zhongming
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING ALPHA-1,3
; TITLE OF INVENTION: FUCOSYLTRANSFERASES AND EXPRESSION SYSTEMS FOR MAKING
AND
; TITLE OF INVENTION: EXPRESSING THEM (amended)
; FILE REFERENCE: 07254-049002
; CURRENT APPLICATION NUMBER: US/09/733,524A
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 09/092,315
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: US 60/048,857
; PRIOR FILING DATE: 1997-06-06
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-733-524A-8

Query Match 59.7%; Score 40; DB 4; Length 454;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 NWWGD 7
|||||
Db 32 NWWGD 36

RESULT 13

US-09-092-315-1
; Sequence 1, Application US/09092315
; Patent No. 6399337
; GENERAL INFORMATION:
; APPLICANT: Taylor, Diane E.
; APPLICANT: Ge, Zhongming
; TITLE OF INVENTION: ALPHA-1, 3-FUCOSYLTRANSFERASE
; FILE REFERENCE: 07254/049001
; CURRENT APPLICATION NUMBER: US/09/092,315
; CURRENT FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: US 60/048,857

; EARLIER FILING DATE: 1997-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 464
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-092-315-1

Query Match 59.7%; Score 40; DB 4; Length 464;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 NWWGD 7
|||
Db 32 NWWGD 36

RESULT 14

US-09-733-524A-1

; Sequence 1, Application US/09733524A
; Patent No. 6534298
; GENERAL INFORMATION:
; APPLICANT: Taylor, Diane E.
; APPLICANT: Ge, Zhongming
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING ALPHA-1,3
; TITLE OF INVENTION: FUCOSYLTRANSFERASES AND EXPRESSION SYSTEMS FOR MAKING
AND
; TITLE OF INVENTION: EXPRESSING THEM (amended)
; FILE REFERENCE: 07254-049002
; CURRENT APPLICATION NUMBER: US/09/733,524A
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 09/092,315
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: US 60/048,857
; PRIOR FILING DATE: 1997-06-06
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 464
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-733-524A-1

Query Match 59.7%; Score 40; DB 4; Length 464;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 NWWGD 7
|||
Db 32 NWWGD 36

RESULT 15

US-09-092-315-5

; Sequence 5, Application US/09092315
; Patent No. 6399337

```

; GENERAL INFORMATION:
; APPLICANT: Taylor, Diane E.
; APPLICANT: Ge, Zhongming
; TITLE OF INVENTION: ALPHA-1, 3-FUCOSYLTRANSFERASE
; FILE REFERENCE: 07254/049001
; CURRENT APPLICATION NUMBER: US/09/092,315
; CURRENT FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: US 60/048,857
; EARLIER FILING DATE: 1997-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 476
;   TYPE: PRT
;   ORGANISM: Helicobacter pylori
US-09-092-315-5

```

```

Query Match          59.7%;  Score 40;  DB 4;  Length 476;
Best Local Similarity 100.0%;  Pred. No. 1e+02;
Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

```

```

Qy          3 NWWGD 7
            |||||
Db          33 NWWGD 37

```

```

Search completed: November 13, 2003, 09:54:55
Job time : 10.6875 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 30.2812 Seconds
(without alignments)
47.176 Million cell updates/sec

Title: US-09-228-866-2
Perfect score: 67
Sequence: 1 CENWWGDVC 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result		%					
No.	Score	Query	Match	Length	DB	ID	Description
1	67	100.0	9	18	AAW13416		Brain homing pepti
2	67	100.0	9	21	AAB07388		Brain homing pepti
3	67	100.0	9	22	AAE11794		Phage peptide #2 t
4	67	100.0	9	23	AAU10705		Brain homing pepti
5	67	100.0	9	24	ABU59530		Brain receptor tar
6	47	70.1	83	22	ABG25219		Novel human diagno
7	47	70.1	252	22	ABB59632		Drosophila melanog
8	44	65.7	61	22	AAU60753		Propionibacterium
9	43	64.2	1050	24	AAE32727		KIAA0032 protein.
10	43	64.2	1050	24	AAE32731		HERC3 protein (var
11	43	64.2	1054	24	AAE32732		HERC3 protein (var
12	42	62.7	79	22	AAU59363		Propionibacterium
13	42	62.7	755	22	AAB94435		Human protein sequ
14	42	62.7	755	23	ABP69413		Human polypeptide
15	42	62.7	1088	23	ABJ05495		Human breast cance
16	42	62.7	1088	23	ABJ01044		Human breast speci
17	41	61.2	40	22	AAM85611		Human immune/haema
18	41	61.2	45	22	AAM06485		Human foetal prote
19	41	61.2	206	20	AAAY00145		Enterococcus faeca
20	41	61.2	206	23	ABP43364		E faecalis EF071 a
21	41	61.2	206	24	ABU13643		Enterococcus faeca
22	41	61.2	207	20	AAAY00144		Enterococcus faeca
23	41	61.2	207	23	ABP43363		E faecalis EF071 p
24	41	61.2	207	24	ABU13642		Enterococcus faeca
25	40.5	60.4	1572	18	AAW27160		Mouse receptor ME2
26	40.5	60.4	2707	18	AAW27161		Mouse receptor ME2
27	40	59.7	20	22	ABB45238		Rabbit albumin-bin
28	40	59.7	146	23	ABU51720		Helicobacter pylor
29	40	59.7	187	22	ABG19166		Novel human diagno
30	40	59.7	189	22	ABG04154		Novel human diagno
31	40	59.7	206	22	ABG04155		Novel human diagno
32	40	59.7	212	22	AAG89892		C glutamicum prote
33	40	59.7	221	22	ABG04153		Novel human diagno
34	40	59.7	243	23	ABG96331		Human ovarian canc
35	40	59.7	243	23	ABG92083		Human receptors an
36	40	59.7	244	22	AAO06091		Human polypeptide
37	40	59.7	248	22	ABG19167		Novel human diagno
38	40	59.7	250	22	AAE09454		Human sbg72825FOLA
39	40	59.7	257	23	ABG96330		Human ovarian canc
40	40	59.7	270	23	ABP41366		Human ovarian anti
41	40	59.7	368	22	ABG19677		Novel human diagno
42	40	59.7	418	23	ABU52257		Helicobacter pylor
43	40	59.7	424	23	ABG30885		H. pylori alpha1,3
44	40	59.7	454	23	ABG30887		H. pylori alpha1,3
45	40	59.7	455	21	AAAY54499		Mouse liver angiop

ALIGNMENTS

RESULT 1

AAW13416

ID AAW13416 standard; Peptide; 9 AA.

XX

AC AAW13416;

XX

DT 15-JAN-1998 (first entry)

XX

DE Brain homing peptide.

XX

KW Brain homing peptide; in vivo panning; screening; phage display;
KW drug delivery.

XX

OS Synthetic.

XX

PN WO9710507-A1.

XX

PD 20-MAR-1997.

XX

PF 10-SEP-1996; 96WO-US14600.

XX

PR 11-SEP-1995; 95US-0526710.

PR 11-SEP-1995; 95US-0526708.

XX

PA (LJOL-) LA JOLLA CANCER RES FOUND.

XX

PI Pasqualini R, Ruoslahti E;

XX

DR WPI; 1997-202359/18.

XX

PT Obtaining compound that homes to selected organ or tissue - by in
PT vivo panning method, specifically to identify brain, kidney,
PT angiogenic vasculature or tumour tissue homing peptide(s)

XX

PS Claim 14; Page 67; 75pp; English.

XX

CC This synthetic peptide is a claimed example of a brain-homing
CC peptide that was identified using a novel method for obtaining
CC molecules that home to a selected organ or tissue. This in vivo
CC panning method typically involves administering a phage display
CC library to a subject, and identifying expressed peptides which
CC home to the desired organ or tissue, e.g. brain, kidney, angiogenic
CC vascular tissue or tumour tissue. The isolated peptides (see
CC AAW13412-52, AAW11181-86) can be used to target e.g. drugs, toxins or
CC labels to the selected organ/tissue (claimed) or to identify and/or
CC isolate target molecules (claimed). The peptides can be directly
CC identified in vivo, as compared to prior art in vitro screening
CC methods, which require further examination to see if they maintain
CC specificity in vivo.

XX

SQ Sequence 9 AA;

Query Match 100.0%; Score 67; DB 18; Length 9;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

1 CENWWGDVC 9

Db

|||||||
1 CENWWGDVC 9

RESULT 2

AAB07388

ID AAB07388 standard; peptide; 9 AA.

XX

AC AAB07388;

XX

DT 17-OCT-2000 (first entry)

XX

DE Brain homing peptide # 2.

XX

KW Brain; homing peptide; organ targeting; tissue targeting; mouse; cyclic.

XX

OS Mus sp.

XX

FH Key Location/Qualifiers

FT Disulfide-bond 1..9

FT /note= "Can optionally form a cyclic peptide"

XX

PN US6068829-A.

XX

PD 30-MAY-2000.

XX

PF 23-JUN-1997; 97US-0862855.

XX

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Pasqualini R, Ruoslahti E;

XX

DR WPI; 2000-410850/35.

XX

PT Identifying and recovering organ homing molecules or peptides by in

PT vivo panning comprises administering a library of diverse peptides

PT linked to a tag which facilitates recovery of these peptides -

XX

PS Example 2; Column 17; 20pp; English.

XX

CC The present sequence is a mouse brain homing peptide. This sequence was

CC identified by using in vivo panning to screen a library of potential

CC organ homing molecules. The present sequence can be used to direct a

CC moiety to a the brain tissue, by linking the moiety to the present

CC sequence. Examples of potential moieties are drugs, toxins or a

CC detectable label.

XX

SQ Sequence 9 AA;

Query Match 100.0%; Score 67; DB 21; Length 9;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 CENWWGDVC 9

Db

|||||||
1 CENWWGDVC 9

RESULT 3

AAE11794

ID AAE11794 standard; peptide; 9 AA.

XX

AC AAE11794;

XX

DT 18-DEC-2001 (first entry)

XX

DE Phage peptide #2 targetted to brain.

XX

KW Enriched library fraction; brain; kidney; tumour; panning; diagnostic;

KW molecular medicine; drug delivery; peptidomimetic; pharmaceutical.

XX

OS Bacteriophage.

XX

PN US6296832-B1.

XX

PD 02-OCT-2001.

XX

PF 08-JAN-1999; 99US-0226985.

XX

PR 23-JUN-1997; 97US-0862855.

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ruoslahti E, Pasqualini R;

XX

DR WPI; 2001-610691/70.

XX

PT Enriched library fraction comprising molecules recovered by in vivo

PT panning that selectively home to a selected organ or tissue useful for

PT treating disease or in diagnostic methods -

XX

PS Example 2; Column 17; 21pp; English.

XX

CC The invention relates to an enriched library fraction containing
CC molecules that selectively home to a selected organ or tissue such as
CC brain, kidney or tumour recovered by in vivo panning. The invention
CC generally relates to the field of molecular medicine, drug delivery and
CC to a method of invivo panning for identifying a molecule that homes to a
CC specific organ. The molecules, e.g., peptides, peptidomimetics, proteins
CC and fragments of proteins contained in an enriched library fraction may
CC be administered to a subject as part of a pharmaceutical composition to
CC treat disease or in diagnostic methods. The present sequence is a
CC peptide from bacteriophage targetted to brain.

XX

SQ Sequence 9 AA;

Query Match 100.0%; Score 67; DB 22; Length 9;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CENWWGDVC 9
|||
Db 1 CENWWGDVC 9

RESULT 4
AAU10705

ID AAU10705 standard; peptide; 9 AA.

XX

AC AAU10705;

XX

DT 12-MAR-2002 (first entry)

XX

DE Brain homing peptide #2 useful for delivery of target molecules.

XX

KW Organ targeting; tissue targeting; cancer; tumour homing molecule;
KW delivery of target molecule; brain homing peptide.

XX

OS Synthetic.

XX

PN US6306365-B1.

XX

PD 23-OCT-2001.

XX

PF 08-JAN-1999; 99US-0227906.

XX

PR 23-JUN-1997; 97US-0862855.

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ruoslahti E, Pasqualini R;

XX

DR WPI; 2002-040196/05.

XX

PT Recovering molecules that home to an organ or tissue, useful for
PT identifying molecules that home to a specific organ or tissue, e.g.
PT identifying a tumour homing molecule to identify the presence of cancer,
PT by in vivo panning of a library -

XX

PS Example 2; Column 17; 21pp; English.

XX

CC The present invention relates to a method of recovering molecules that
CC home to a selected organ or tissue. The method comprises administering
CC to the subject the library of diverse molecules, collecting a sample of
CC the selected organ or tissue (e.g. brain or kidney), and recovering from
CC the sample several molecules that home to the selected organ or tissue.
CC The method is useful for identifying molecules, particularly useful for
CC screening large number of molecules (e.g. peptides), that home to a
CC specific organ. The identified molecule is useful for e.g. raising an
CC antibody specific for a target molecule, targeting a desired moiety
CC (e.g. drug, toxin or detectable label) to the selected organ.
CC Specifically, the method is useful for identifying the presence of cancer
CC in a subject by linking an appropriate moiety to a tumour homing
CC molecule. The present method provides a direct means for identifying

CC molecules that specifically home to a selected organ and, therefore
CC provides a significant advantage over previous methods, which require
CC that a molecule identified using an in vitro screening method
CC subsequently be examined to determine if it maintains its specificity in
CC vivo. AAU10704-AAU10723 represent brain homing peptides described in
CC the present invention.

XX

SQ Sequence 9 AA;

Query Match 100.0%; Score 67; DB 23; Length 9;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CENWWGDVC 9

|||||||

Db 1 CENWWGDVC 9

RESULT 5

ABU59530

ID ABU59530 standard; Peptide; 9 AA.

XX

AC ABU59530;

XX

DT 22-APR-2003 (first entry)

XX

DE Brain receptor targeting peptide #2.

XX

KW Targeting ligand; bioactive agent; polymer matrix; cancer; cytostatic;

KW cathepsin-D substrate; peptides; neuroreceptor; adrenal receptor;

KW fibronectin; vitronectin; integrin; RGD motif; angiogenic endothelium;

KW tumour; cationic cancer-targeting peptide.

XX

OS Synthetic.

XX

PN US2002041898-A1.

XX

PD 11-APR-2002.

XX

PF 25-JUL-2001; 2001US-0912609.

XX

PR 05-JAN-2000; 2000US-0478124.

PR 31-OCT-2000; 2000US-0703474.

XX

PA (UNGE/) UNGER E C.

PA (MATS/) MATSUNAGA T O.

PA (RAMA/) RAMASWAMI V.

PA (ROMA/) ROMANOWSKI M J.

XX

PI Unger EC, Matsunaga TO, Ramaswami V, Romanowski MJ;

XX

DR WPI; 2003-208921/20.

XX

PT Targeted delivery system comprising a bioactive agent homogeneously

PT dispersed in a targeted matrix is especially useful in cancer therapy

PT -

XX

PS Claim 23; Page 37; 46pp; English.

XX

CC The invention relates to a composition comprising a bioactive agent
CC homogeneously dispersed in a targeted matrix (polymer and targeting
CC ligand). Also included are a targeted matrix for use as a delivery
CC vehicle comprising a polymer associated with a targeting ligand,
CC enhancing the bioavailability of an agent comprising administration
CC of the composition and treating cancer comprising administration of the
CC novel composition. The method is useful for targeted delivery of a drug,
CC especially in cancer therapy. The targeting ligand may be a peptide.
CC Examples of targeting peptides are disclosed including cathepsin-D
CC substrate peptides, peptides targeting receptors in the brain and
CC kidney, peptides recognising fibronectin- and vitronectin-binding
CC integrins, peptides targeting the RGD (Arg-Gly-Asp)-motif in, e.g.,
CC antibodies, peptides targeting the angiogenic endothelium of solid
CC tumours, tissue specific peptides (e.g. of lung, skin, pancreas,
CC intestine, uterus, adrenal gland and retina), and cationic cancer-
CC targeting peptides. The present sequence is a peptide targeting
CC ligand disclosed in the invention.

XX

SQ Sequence 9 AA;

Query Match 100.0%; Score 67; DB 24; Length 9;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CENWWGDVC 9
| | | | | | | |
Db 1 CENWWGDVC 9

RESULT 6

ABG25219

ID ABG25219 standard; Protein; 83 AA.

XX

AC ABG25219;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #25210.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

DR N-PSDB; AAS89406.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX

PS Claim 20; SEQ ID No 55578; 103pp; English.

XX

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 83 AA;

Query Match 70.1%; Score 47; DB 22; Length 83;

Best Local Similarity 85.7%; Pred. No. 7.9;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 NWWGDVC 9

|||||

Db 33 NWWGSVC 39

RESULT 7

ABB59632

ID ABB59632 standard; Protein; 252 AA.

XX

AC ABB59632;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 5688.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 DR N-PSDB; ABL03735.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 5688; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 252 AA;

Query Match 70.1%; Score 47; DB 22; Length 252;
 Best Local Similarity 75.0%; Pred. No. 23;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ENWWGDVC 9
 |||| :||
 Db 138 ENWWANVC 145

RESULT 8
 AAU60753
 ID AAU60753 standard; Protein; 61 AA.
 XX
 AC AAU60753;
 XX
 DT 27-FEB-2002 (first entry)
 XX
 DE Propionibacterium acnes immunogenic protein #21649.

XX
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.
 XX
 OS *Propionibacterium acnes*.
 XX
 PN WO200181581-A2.
 XX
 PD 01-NOV-2001.
 XX
 PF 20-APR-2001; 2001WO-US12865.
 XX
 PR 21-APR-2000; 2000US-199047P.
 PR 02-JUN-2000; 2000US-208841P.
 PR 07-JUL-2000; 2000US-216747P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
 XX
 DR WPI; 2001-616774/71.
 DR N-PSDB; AAS59612.
 XX
 PT *Propionibacterium acnes* polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris -
 XX
 PS Example 1; SEQ ID No 21948; 1069pp; English.
 XX
 CC Sequences AAU39105-AAU68017 represent *Propionibacterium acnes* immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC *P. acnes*. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC *P. acnes* is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of *P. acnes* in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for *P. acnes* proteins. These antibodies can be used to
 CC downregulate expression and activity of *P. acnes* polypeptides and
 CC therefore treat *P. acnes* infections. The antibodies may also be used as
 CC diagnostic agents for determining *P. acnes* presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at [ftp.wipo.int/pub/published_pct_sequences](ftp:wipo.int/pub/published_pct_sequences).
 XX
 SQ Sequence 61 AA;

Query Match 65.7%; Score 44; DB 22; Length 61;
 Best Local Similarity 66.7%; Pred. No. 16;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CENWWGDVC 9
| | : | | |
Db 49 CANYWQDVC 57

RESULT 9

AAE32727

ID AAE32727 standard; Protein; 1050 AA.

XX

AC AAE32727;

XX

DT 24-MAR-2003 (first entry)

XX

DE KIAA0032 protein.

XX

KW Viral infection; lymphosarcoma; human immunodeficiency virus; hepatitis;
KW poliomyelitis; HIV; measles; protein therapy; KIAA0032 protein.

XX

OS Unidentified.

XX

FH Key Location/Qualifiers

FT Domain 52..102

FT /note= "RCC1 domain"

FT Domain 1012..1047

FT /note= "HECT domain"

XX

PN WO200290549-A2.

XX

PD 14-NOV-2002.

XX

PF 12-MAR-2002; 2002WO-IB02106.

XX

PR 12-MAR-2001; 2001US-275224P.

PR 31-JUL-2001; 2001US-308958P.

PR 07-DEC-2001; 2001US-340170P.

XX

PA (PROT-) PROTEOLOGICS LTD.

XX

PI Greener T, Moskowitz H, Reiss Y, Alroy I;

XX

DR WPI; 2003-111976/10.

DR N-PSDB; AAD50461.

XX

PT New protein complex comprising HECT-RCC1, viral maturation scaffolding
PT protein (VMSP), and/or HIV gag protein, useful for treating viral
PT infections, such lymphosarcoma, HIV, hepatitis, poliomyelitis, measles,
PT or Ebola -

XX

PS Disclosure; Fig 24; 150pp; English.

XX

CC The invention relates to a method for modulation of viral maturation.

CC The invention also provides an isolated protein complex comprising a

CC HECT-RCC1 selected from HETT-WW, HECT-RCC1, Gag protein, Gag late

CC domain, P13, actin, myosin, Hsp60, Hsp70, Hsp90, STAM1, STAM2A, STAM2B,

CC VHS-UIM, GTPase, E2 enzyme, tsg101, cullin, HERC1, HERC2, HERC3, Nedd4

CC -like protein or clathrin. The complexes, proteins, antibodies and
CC methods are useful for treating viral infections, such as lymphosarcoma,
CC human immunodeficiency virus (HIV), hepatitis, poliomyelitis, measles,
CC or Ebola and for inhibiting budding in a subject. They are also useful
CC in diagnostic assays for determining whether a cell is infected with a
CC virus and for characterising the nature, progression and/or infectivity
CC of the infection. The invention is also useful in protein therapy. The
CC present sequence is KIAA0032 protein used to illustrate the method of
CC the invention.

XX

SQ Sequence 1050 AA;

Query Match 64.2%; Score 43; DB 24; Length 1050;
Best Local Similarity 62.5%; Pred. No. 3.5e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ENWWGDVC 9
: ||| ||
Db 545 DNWWSQVC 552

RESULT 10

AAE32731

ID AAE32731 standard; Protein; 1050 AA.

XX

AC AAE32731;

XX

DT 24-MAR-2003 (first entry)

XX

DE HERC3 protein (var1).

XX

KW Viral infection; lymphosarcoma; human immunodeficiency virus; hepatitis;
KW poliomyelitis; HIV; measles; protein therapy; HERC3 protein.

XX

OS Unidentified.

XX

PN WO200290549-A2.

XX

PD 14-NOV-2002.

XX

PF 12-MAR-2002; 2002WO-IB02106.

XX

PR 12-MAR-2001; 2001US-275224P.

PR

PR 31-JUL-2001; 2001US-308958P.

PR

PR 07-DEC-2001; 2001US-340170P.

XX

PA (PROT-) PROTEOLOGICS LTD.

XX

PI Greener T, Moskowitz H, Reiss Y, Alroy I;

XX

DR WPI; 2003-111976/10.

DR

DR N-PSDB; AAD50465.

XX

PT New protein complex comprising HECT-RCCL1, viral maturation scaffolding

PT protein (VMSP), and/or HIV gag protein, useful for treating viral

PT infections, such lymphosarcoma, HIV, hepatitis, poliomyelitis, measles,

PT or Ebola -

XX
PS Claim 36; Fig 28; 150pp; English.

XX
CC The invention relates to a method for modulation of viral maturation.
CC The invention also provides an isolated protein complex comprising a
CC HECT-RCC1 selected from HETT-WW, HECT-RCC1, Gag protein, Gag late
CC domain, P13, actin, myosin, Hsp60, Hsp70, Hsp90, STAM1, STAM2A, STAM2B,
CC VHS-UIIM, GTPase, E2 enzyme, tsg101, cullin, HERC1, HERC2, HERC3, Nedd4
CC -like protein or clathrin. The complexes, proteins, antibodies and
CC methods are useful for treating viral infections, such as lymphosarcoma,
CC human immunodeficiency virus (HIV), hepatitis, poliomyelitis, measles,
CC or Ebola and for inhibiting budding in a subject. They are also useful
CC in diagnostic assays for determining whether a cell is infected with a
CC virus and for characterising the nature, progression and/or infectivity
CC of the infection. The invention is also useful in protein therapy. The
CC present sequence is HERC3 protein used to illustrate the method of the
CC invention.

XX
SQ Sequence 1050 AA;

Query Match 64.2%; Score 43; DB 24; Length 1050;
Best Local Similarity 62.5%; Pred. No. 3.5e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ENWWGDVC 9
: ||| ||
Db 545 DNWWSQVC 552

RESULT 11
AAE32732
ID AAE32732 standard; Protein; 1054 AA.
XX
AC AAE32732;
XX
DT 24-MAR-2003 (first entry)
XX
DE HERC3 protein (var2).
XX
KW Viral infection; lymphosarcoma; human immunodeficiency virus; hepatitis;
KW poliomyelitis; HIV; measles; protein therapy; HERC3 protein.
XX
OS Unidentified.
XX
PN WO200290549-A2.
XX
PD 14-NOV-2002.
XX
PF 12-MAR-2002; 2002WO-IB02106.
XX
PR 12-MAR-2001; 2001US-275224P.
PR 31-JUL-2001; 2001US-308958P.
PR 07-DEC-2001; 2001US-340170P.
XX
PA (PROT-) PROTEOLOGICS LTD.
XX
PI Greener T, Moskowitz H, Reiss Y, Alroy I;

XX
 DR WPI; 2003-111976/10.
 DR N-PSDB; AAD50466.
 XX
 PT New protein complex comprising HECT-RCC1, viral maturation scaffolding
 PT protein (VMSP), and/or HIV gag protein, useful for treating viral
 PT infections, such lymphosarcoma, HIV, hepatitis, poliomyelitis, measles,
 PT or Ebola -
 XX
 PS Claim 36; Fig 29; 150pp; English.
 XX
 CC The invention relates to a method for modulation of viral maturation.
 CC The invention also provides an isolated protein complex comprising a
 CC HECT-RCC1 selected from HETT-WW, HECT-RCC1, Gag protein, Gag late
 CC domain, P13, actin, myosin, Hsp60, Hsp70, Hsp90, STAM1, STAM2A, STAM2B,
 CC VHS-UIM, GTPase, E2 enzyme, tsg101, cullin, HERC1, HERC2, HERC3, Nedd4
 CC -like protein or clathrin. The complexes, proteins, antibodies and
 CC methods are useful for treating viral infections, such as lymphosarcoma,
 CC human immunodeficiency virus (HIV), hepatitis, poliomyelitis, measles,
 CC or Ebola and for inhibiting budding in a subject. They are also useful
 CC in diagnostic assays for determining whether a cell is infected with a
 CC virus and for characterising the nature, progression and/or infectivity
 CC of the infection. The invention is also useful in protein therapy. The
 CC present sequence is HERC3 protein used to illustrate the method of the
 CC invention.
 XX
 SQ Sequence 1054 AA;

Query Match 64.2%; Score 43; DB 24; Length 1054;
 Best Local Similarity 62.5%; Pred. No. 3.5e+02;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ENWWGDVC 9
 :||| ||
 Db 549 DNWWSQVC 556

RESULT 12
 AAU59363
 ID AAU59363 standard; Protein; 79 AA.
 XX
 AC AAU59363;
 XX
 DT 27-FEB-2002 (first entry)
 XX
 DE Propionibacterium acnes immunogenic protein #20259.
 XX
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.
 XX
 OS Propionibacterium acnes.
 XX
 PN WO200181581-A2.
 XX
 PD 01-NOV-2001.

XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59602.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1; SEQ ID No 20558; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 79 AA;

Query Match 62.7%; Score 42; DB 22; Length 79;
Best Local Similarity 66.7%; Pred. No. 41;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CENWWGDVC 9
||: |||
Db 43 CESAWS DVC 51

RESULT 13
AAB94435
ID AAB94435 standard; Protein; 755 AA.
XX

AC AAB94435;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:15056.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 15056; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

XX

SQ Sequence 755 AA;

Query Match 62.7%; Score 42; DB 22; Length 755;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CENWW 5
| | | |
Db 327 CENWW 331

RESULT 14

ABP69413

ID ABP69413 standard; Protein; 755 AA.

XX

AC ABP69413;

XX

DT 20-JAN-2003 (first entry)

XX

DE Human polypeptide SEQ ID NO 1460.

XX

KW Human; genome mapping; gene therapy; food supplement; virus; fungus;
KW cell-proliferative disorder; neurodegenerative disease; bacterial;
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
KW haemostatic; vulnerary; fungicide; antibacterial; virucide; protozoacide;
KW antiarthritic.

XX

OS Homo sapiens.

XX

PN WO200270539-A2.

XX

PD 12-SEP-2002.

XX

PF 05-MAR-2002; 2002WO-US05095.

XX

PR 05-MAR-2001; 2001US-0799451.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
PI Wehrman T, Wang J, Wang D, Drmanac RT;

XX

DR WPI; 2002-759812/82.

DR N-PSDB; ABZ11630.

XX

PT New polynucleotides comprising sequences assembled from expressed
PT sequence tags (ESTs), useful for treating cell-proliferative,
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or
PT platelet or coagulation disorders -

XX

PS Claim 9; SEQ ID NO 1460; 1012pp + Sequence Listing; English.

XX
CC The invention relates to an isolated polynucleotide (I) comprising a
CC nucleotide sequence selected from any of 948 sequences
CC (ABZ11119-ABZ12066) or their mature protein coding portion, active domain
CC coding protein or complementary sequences. The polynucleotides are useful
CC for identifying expressed genes or for physical mapping of human genome.
CC The encoded polypeptides (ABP68902-ABP69849) are useful as molecular
CC weight markers, as a food supplement, for generating antibodies, in
CC medical imaging, screening and diagnostic assays and for treating
CC cell-proliferative disorders (cancer), neurodegenerative diseases
CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple
CC sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid
CC disorders, platelet or coagulation disorders, wound, burns, incision,
CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,
CC parasitic), arthritis, etc.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 755 AA;

Query Match 62.7%; Score 42; DB 23; Length 755;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CENWW 5
| | | |
Db 327 CENWW 331

RESULT 15

ABJ05495

ID ABJ05495 standard; Protein; 1088 AA.

XX

AC ABJ05495;

XX

DT 14-NOV-2002 (first entry)

XX

DE Human breast cancer associated polypeptide SEQ ID NO: 255.

XX

KW Human; breast specific gene; breast specific protein; breast cancer;
KW gene therapy; cytostatic.

XX

OS Homo sapiens.

XX

PN WO200264611-A1.

XX

PD 22-AUG-2002.

XX

PF 12-FEB-2002; 2002WO-US04197.

XX

PR 13-FEB-2001; 2001US-268292P.

XX

PA (DIAD-) DIADEXUS INC.

XX

PI Salceda S, Macina RA, Hu P, Recipon H, Karra K, Cafferkey R;
PI Sun Y, Liu C;

XX
DR WPI; 2002-657582/70.
XX
PT New breast specific nucleic acids and proteins, useful for identifying,
PT diagnosing, monitoring, staging, imaging, and treating breast cancer
PT and non-cancerous disease states in breast tissue, and in gene therapy
PT -
XX
PS Claim 11; Page 330-334; 367pp; English.
XX
CC The present invention provides human breast specific coding sequences and
CC proteins. These can be used in the diagnosis and treatment of breast
CC cancer and non-cancerous diseases of the breast. The present sequence is
CC a polypeptide of the invention.
XX
SQ Sequence 1088 AA;

Query Match 62.7%; Score 42; DB 23; Length 1088;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CENWW 5
|||
Db 660 CENWW 664

Search completed: November 13, 2003, 09:45:22
Job time : 31.2812 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:45:35 ; Search time 18.6562 Seconds
(without alignments)
88.069 Million cell updates/sec

Title: US-09-228-866-2
Perfect score: 67
Sequence: 1 CENWWGDVC 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43	64.2	1050	15	US-10-097-534-24	Sequence 24, Appl
2	43	64.2	1050	15	US-10-097-534-28	Sequence 28, Appl
3	43	64.2	1054	15	US-10-097-534-29	Sequence 29, Appl
4	42	62.7	1088	14	US-10-001-887-127	Sequence 127, App
5	42	62.7	1088	15	US-10-074-475-255	Sequence 255, App
6	40	59.7	212	10	US-09-738-626-3646	Sequence 3646, Ap
7	40	59.7	243	15	US-10-097-340-117	Sequence 117, App
8	40	59.7	250	12	US-10-203-708-41	Sequence 41, Appl
9	40	59.7	257	15	US-10-097-340-111	Sequence 111, App
10	40	59.7	372	12	US-10-189-977-13	Sequence 13, Appl
11	40	59.7	372	14	US-10-120-319-13	Sequence 13, Appl
12	40	59.7	424	9	US-09-733-524-16	Sequence 16, Appl
13	40	59.7	425	12	US-10-189-977-6	Sequence 6, Appli
14	40	59.7	425	12	US-10-392-098-6	Sequence 6, Appli
15	40	59.7	425	14	US-10-120-319-6	Sequence 6, Appli
16	40	59.7	454	9	US-09-733-524-18	Sequence 18, Appl
17	40	59.7	454	12	US-10-189-977-8	Sequence 8, Appli
18	40	59.7	454	12	US-10-392-098-8	Sequence 8, Appli
19	40	59.7	454	14	US-10-120-319-8	Sequence 8, Appli
20	40	59.7	455	12	US-10-460-125-2	Sequence 2, Appli
21	40	59.7	464	12	US-10-189-977-1	Sequence 1, Appli
22	40	59.7	464	12	US-10-392-098-1	Sequence 1, Appli
23	40	59.7	464	14	US-10-120-319-1	Sequence 1, Appli
24	40	59.7	476	9	US-09-733-524-15	Sequence 15, Appl
25	40	59.7	476	12	US-10-189-977-5	Sequence 5, Appli
26	40	59.7	476	12	US-10-392-098-5	Sequence 5, Appli
27	40	59.7	476	14	US-10-120-319-5	Sequence 5, Appli
28	40	59.7	478	12	US-10-189-977-7	Sequence 7, Appli

29	40	59.7	478	12	US-10-392-098-7	Sequence 7, Appli
30	40	59.7	478	14	US-10-120-319-7	Sequence 7, Appli
31	40	59.7	479	9	US-09-733-524-17	Sequence 17, Appl
32	40	59.7	485	9	US-09-733-524-2	Sequence 2, Appli
33	40	59.7	486	12	US-10-189-977-2	Sequence 2, Appli
34	40	59.7	486	12	US-10-392-098-2	Sequence 2, Appli
35	40	59.7	486	14	US-10-120-319-2	Sequence 2, Appli
36	40	59.7	501	9	US-09-733-524-1	Sequence 1, Appli
37	39	58.2	12	12	US-10-190-082-556	Sequence 556, App
38	39	58.2	237	10	US-09-738-626-6721	Sequence 6721, Ap
39	39	58.2	285	10	US-09-738-626-4153	Sequence 4153, Ap
40	39	58.2	500	10	US-09-323-998D-59	Sequence 59, Appl
41	39	58.2	790	15	US-10-156-761-9515	Sequence 9515, Ap
42	39	58.2	832	12	US-10-369-294-21	Sequence 21, Appl
43	39	58.2	853	12	US-10-369-294-13	Sequence 13, Appl
44	38	56.7	86	10	US-09-751-100B-52	Sequence 52, Appl
45	38	56.7	86	10	US-09-751-100B-53	Sequence 53, Appl

ALIGNMENTS

RESULT 1

US-10-097-534-24

; Sequence 24, Application US/10097534

; Publication No. US20030049607A1

; GENERAL INFORMATION:

; APPLICANT: GREENER, TSVIKA

; APPLICANT: MOSKOWITZ, HAIM

; APPLICANT: REISS, YUVAL

; APPLICANT: ALROY, IRIS

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE MODULATION OF VIRAL

; TITLE OF INVENTION: MATURATION

; FILE REFERENCE: PLV-001.01

; CURRENT APPLICATION NUMBER: US/10/097,534

; CURRENT FILING DATE: 2002-03-12

; PRIOR APPLICATION NUMBER: 60/275,224

; PRIOR FILING DATE: 2001-03-12

; PRIOR APPLICATION NUMBER: 60/308,958

; PRIOR FILING DATE: 2001-07-31

; PRIOR APPLICATION NUMBER: 60/340,170

; PRIOR FILING DATE: 2001-12-07

; NUMBER OF SEQ ID NOS: 71

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 24

; LENGTH: 1050

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-097-534-24

Query Match 64.2%; Score 43; DB 15; Length 1050;

Best Local Similarity 62.5%; Pred. No. 3.7e+02;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ENWWGDVC 9

:||| ||

Db 545 DNWWSQVC 552

RESULT 2

US-10-097-534-28

; Sequence 28, Application US/10097534
; Publication No. US20030049607A1
; GENERAL INFORMATION:
; APPLICANT: GREENER, TSVIKA
; APPLICANT: MOSKOWITZ, HAIM
; APPLICANT: REISS, YUVAL
; APPLICANT: ALROY, IRIS
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE MODULATION OF VIRAL
; TITLE OF INVENTION: MATURATION
; FILE REFERENCE: PLV-001.01
; CURRENT APPLICATION NUMBER: US/10/097,534
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/275,224
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/308,958
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/340,170
; PRIOR FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 1050
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-534-28

Query Match 64.2%; Score 43; DB 15; Length 1050;
Best Local Similarity 62.5%; Pred. No. 3.7e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ENWWGDVC 9
: ||| ||
Db 545 DNWWSQVC 552

RESULT 3

US-10-097-534-29

; Sequence 29, Application US/10097534
; Publication No. US20030049607A1
; GENERAL INFORMATION:
; APPLICANT: GREENER, TSVIKA
; APPLICANT: MOSKOWITZ, HAIM
; APPLICANT: REISS, YUVAL
; APPLICANT: ALROY, IRIS
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE MODULATION OF VIRAL
; TITLE OF INVENTION: MATURATION
; FILE REFERENCE: PLV-001.01
; CURRENT APPLICATION NUMBER: US/10/097,534
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/275,224
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/308,958
; PRIOR FILING DATE: 2001-07-31

; PRIOR APPLICATION NUMBER: 60/340,170
; PRIOR FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 1054
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-534-29

Query Match 64.2%; Score 43; DB 15; Length 1054;
Best Local Similarity 62.5%; Pred. No. 3.7e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ENWWGDVC 9
: ||| ||
Db 549 DNWWSQVC 556

RESULT 4

US-10-001-887-127
; Sequence 127, Application US/10001887
; Publication No. US20020155464A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Cafferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
Genes and Proteins
; FILE REFERENCE: DEX-0269
; CURRENT APPLICATION NUMBER: US/10/001,887
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,998
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: 60/252,563
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 127
; LENGTH: 1088
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-887-127

Query Match 62.7%; Score 42; DB 14; Length 1088;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CENWW 5
|||||
Db 660 CENWW 664

RESULT 5

US-10-074-475-255
 ; Sequence 255, Application US/10074475
 ; Publication No. US20030092898A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Salceda, Susana
 ; APPLICANT: Macina, Roberto
 ; APPLICANT: Hu, Ping
 ; APPLICANT: Recipon, Herve
 ; APPLICANT: Karra, Kalpana
 ; APPLICANT: Cafferkey, Robert
 ; APPLICANT: Sun, Yongming
 ; APPLICANT: Liu, Chenghua
 ; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
 ; TITLE OF INVENTION: Genes and Proteins
 ; FILE REFERENCE: DEX-0313
 ; CURRENT APPLICATION NUMBER: US/10/074,475
 ; CURRENT FILING DATE: 2002-02-13
 ; PRIOR APPLICATION NUMBER: 60/268,292
 ; PRIOR FILING DATE: 2001-02-13
 ; NUMBER OF SEQ ID NOS: 295
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 255
 ; LENGTH: 1088
 ; TYPE: PRT
 ; ORGANISM: Homo sapien
 US-10-074-475-255

Query Match 62.7%; Score 42; DB 15; Length 1088;
 Best Local Similarity 100.0%; Pred. No. 5.1e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CENWW 5
 |||||
 Db 660 CENWW 664

RESULT 6
 US-09-738-626-3646
 ; Sequence 3646, Application US/09738626
 ; Publication No. US20020197605A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NAKAGAWA, SATOSHI
 ; APPLICANT: MIZOGUCHI, HIROSHI
 ; APPLICANT: ANDO, SEIKO
 ; APPLICANT: HAYASHI, MIKIRO
 ; APPLICANT: OCHIAI, KEIKO
 ; APPLICANT: YOKOI, HARUHIKO
 ; APPLICANT: TATEISHI, NAOKO
 ; APPLICANT: SENOH, AKIHIRO
 ; APPLICANT: IKEDA, MASATO
 ; APPLICANT: OZAKI, AKIO
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-125
 ; CURRENT APPLICATION NUMBER: US/09/738,626
 ; CURRENT FILING DATE: 2000-12-18
 ; PRIOR APPLICATION NUMBER: JP 99/377484
 ; PRIOR FILING DATE: 1999-12-16

; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3646
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3646

Query Match 59.7%; Score 40; DB 10; Length 212;
Best Local Similarity 55.6%; Pred. No. 2.6e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CENWWGDVC 9
|:| || :|
Db 35 CDNTWGRLC 43

RESULT 7

US-10-097-340-117

; Sequence 117, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVARAPU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The
Identification,
; TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732

; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 117
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-340-117

Query Match 59.7%; Score 40; DB 15; Length 243;
Best Local Similarity 71.4%; Pred. No. 2.9e+02;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CENWWGD 7
|| || |
Db 137 CERWWED 143

RESULT 8

US-10-203-708-41
; Sequence 41, Application US/10203708
; Publication No. US20030149238A1
; GENERAL INFORMATION:
; APPLICANT: SMITHKLINE BEECHAM CORPORATION
; APPLICANT: SMITHKLINE BEECHAM p.l.c.
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50013
; CURRENT APPLICATION NUMBER: US/10/203,708
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/04703
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/182,172
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 60/186,084
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 41
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-203-708-41

Query Match 59.7%; Score 40; DB 12; Length 250;
Best Local Similarity 71.4%; Pred. No. 3e+02;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CENWWGD 7
|| || |
Db 136 CEEWWED 142

RESULT 9

US-10-097-340-111
 ; Sequence 111, Application US/10097340
 ; Publication No. US20030087250A1
 ; GENERAL INFORMATION:
 ; APPLICANT: John MONAHAN
 ; APPLICANT: Manjula GANNAVARAPU
 ; APPLICANT: Sebastian HOERSCH
 ; APPLICANT: Shubhangi KAMATKAR
 ; APPLICANT: Steve G. KOVATS
 ; APPLICANT: Rachel E. MEYERS
 ; APPLICANT: Michael MORRISEY
 ; APPLICANT: Peter OLANDT
 ; APPLICANT: Ami SEN
 ; APPLICANT: Peter VEIBY
 ; APPLICANT: Gordon B. MILLS
 ; APPLICANT: Robert C. BAST, Jr.
 ; APPLICANT: Karen LU
 ; APPLICANT: Rosemarie SCHMANDT
 ; APPLICANT: Xumei ZHAO
 ; APPLICANT: Karen GLATT
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The
 Identification,
 ; TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer
 ; FILE REFERENCE: MRI-030
 ; CURRENT APPLICATION NUMBER: US/10/097,340
 ; CURRENT FILING DATE: 2002-03-14
 ; PRIOR APPLICATION NUMBER: 60/276,025
 ; PRIOR FILING DATE: 2001-03-14
 ; PRIOR APPLICATION NUMBER: 60/325,149
 ; PRIOR FILING DATE: 2001-09-26
 ; PRIOR APPLICATION NUMBER: 60/276,026
 ; PRIOR FILING DATE: 2001-03-14
 ; PRIOR APPLICATION NUMBER: 60/324,967
 ; PRIOR FILING DATE: 2001/09/26
 ; PRIOR APPLICATION NUMBER: 60/311,732
 ; PRIOR FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: 60/325,102
 ; PRIOR FILING DATE: 2001-09-26
 ; PRIOR APPLICATION NUMBER: 60/323,580
 ; PRIOR FILING DATE: 2001-09-19
 ; NUMBER OF SEQ ID NOS: 363
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 111
 ; LENGTH: 257
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-097-340-111

Query Match 59.7%; Score 40; DB 15; Length 257;
 Best Local Similarity 71.4%; Pred. No. 3e+02;
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CENWWGD 7
 || || |
 Db 139 CEQWWED 145

RESULT 10

US-10-189-977-13

```
; Sequence 13, Application US/10189977
; Publication No. US20030166211A1
; GENERAL INFORMATION:
; APPLICANT: Taylor, Diane E.
; APPLICANT: Ge, Zhongming
; TITLE OF INVENTION: ALPHA-1, 3-FUCOSYLTRANSFERASE
; FILE REFERENCE: 07254/049001
; CURRENT APPLICATION NUMBER: US/10/189,977
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US/09/092,315
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: US 60/048,857
; PRIOR FILING DATE: 1997-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-10-189-977-13
```

```
Query Match          59.7%; Score 40; DB 12; Length 372;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches      5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      3 NWWGD 7
        |||||
Db      32 NWWGD 36
```

RESULT 11

US-10-120-319-13

```
; Sequence 13, Application US/10120319
; Publication No. US20020164749A1
; GENERAL INFORMATION:
; APPLICANT: Taylor, Diane E.
; APPLICANT: Ge, Zhongming
; TITLE OF INVENTION: ALPHA-1, 3-FUCOSYLTRANSFERASE
; FILE REFERENCE: 07254/049001
; CURRENT APPLICATION NUMBER: US/10/120,319
; CURRENT FILING DATE: 2002-04-09
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/092,315
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/048,857
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-10-120-319-13
```

```
Query Match          59.7%; Score 40; DB 14; Length 372;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
```

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 NWWGD 7
 |||||
 Db 32 NWWGD 36

RESULT 12

US-09-733-524-16
 ; Sequence 16, Application US/09733524
 ; Patent No. US20020068347A1
 ; GENERAL INFORMATION:
 ; APPLICANT: The Governors of the University of Alberta, a Canada Corporation
 ; APPLICANT: Taylor, Diane E.
 ; APPLICANT: Ge, Zhongming
 ; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING ALPHA-1,3
 ; TITLE OF INVENTION: FUCOSYLTRANSFERASES AND EXPRESSION SYSTEMS FOR MAKING
 AND
 ; TITLE OF INVENTION: EXPRESSING THEM
 ; FILE REFERENCE: 07254/049002
 ; CURRENT APPLICATION NUMBER: US/09/733,524
 ; CURRENT FILING DATE: 2000-12-14
 ; PRIOR APPLICATION NUMBER: 09/092,315
 ; PRIOR FILING DATE: 1998-06-05
 ; PRIOR APPLICATION NUMBER: 60/048,857
 ; PRIOR FILING DATE: 1997-06-06
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 16
 ; LENGTH: 424
 ; TYPE: PRT
 ; ORGANISM: Helicobacter pylori fucosyltransferase
 ; FEATURE:
 ; NAME/KEY: PEPTIDE
 ; LOCATION: (0)...(0)
 ; OTHER INFORMATION: Strain 26695B
 US-09-733-524-16

Query Match 59.7%; Score 40; DB 9; Length 424;
 Best Local Similarity 100.0%; Pred. No. 4.5e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 NWWGD 7
 |||||
 Db 33 NWWGD 37

RESULT 13

US-10-189-977-6
 ; Sequence 6, Application US/10189977
 ; Publication No. US20030166211A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Taylor, Diane E.
 ; APPLICANT: Ge, Zhongming
 ; TITLE OF INVENTION: ALPHA-1, 3-FUCOSYLTRANSFERASE
 ; FILE REFERENCE: 07254/049001
 ; CURRENT APPLICATION NUMBER: US/10/189,977

; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US/09/092,315
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: US 60/048,857
; PRIOR FILING DATE: 1997-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 425
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-10-189-977-6

Query Match 59.7%; Score 40; DB 12; Length 425;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 NWWGD 7
|||||
Db 33 NWWGD 37

RESULT 14

US-10-392-098-6

; Sequence 6, Application US/10392098
; Publication No. US20030166212A1
; GENERAL INFORMATION:
; APPLICANT: Taylor, Diane E.
; APPLICANT: Ge, Zhongming
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING ALPHA-1,3
; TITLE OF INVENTION: FUCOSYLTRANSFERASES AND EXPRESSION SYSTEMS FOR MAKING
AND
; TITLE OF INVENTION: EXPRESSING THEM (amended)
; FILE REFERENCE: 07254-049002
; CURRENT APPLICATION NUMBER: US/10/392,098
; CURRENT FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: US/09/733,524A
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 09/092,315
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: US 60/048,857
; PRIOR FILING DATE: 1997-06-06
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 425
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-10-392-098-6

Query Match 59.7%; Score 40; DB 12; Length 425;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 NWWGD 7
|||||
Db 33 NWWGD 37

RESULT 15

US-10-120-319-6

; Sequence 6, Application US/10120319
; Publication No. US20020164749A1
; GENERAL INFORMATION:
; APPLICANT: Taylor, Diane E.
; APPLICANT: Ge, Zhongming
; TITLE OF INVENTION: ALPHA-1, 3-FUCOSYLTRANSFERASE
; FILE REFERENCE: 07254/049001
; CURRENT APPLICATION NUMBER: US/10/120,319
; CURRENT FILING DATE: 2002-04-09
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/092,315
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/048,857
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 425
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-10-120-319-6

Query Match 59.7%; Score 40; DB 14; Length 425;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 NWWGD 7
|||
Db 33 NWWGD 37

Search completed: November 13, 2003, 09:58:27
Job time : 18.6562 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:38:30 ; Search time 9.375 Seconds
(without alignments)
92.322 Million cell updates/sec

Title: US-09-228-866-2
Perfect score: 67
Sequence: 1 CENWWGDVC 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	65.7	267	2	T16677	hypothetical prote
2	44	65.7	681	2	S73550	DNA polymerase III
3	44	65.7	1132	2	AF2481	hypothetical prote
4	43	64.2	1054	2	B38919	hypothetical prote
5	41	61.2	242	2	A45724	pectate lyase (EC
6	41	61.2	443	2	T24191	hypothetical prote
7	40	59.7	243	2	A53506	folate receptor ty
8	40	59.7	257	2	A45753	folate-binding pro
9	40	59.7	425	2	C64567	fucosyltransferase
10	40	59.7	476	2	C64601	fucosyltransferase
11	40	59.7	582	2	D72075	hypothetical prote
12	40	59.7	629	2	E86547	hypothetical prote
13	40	59.7	629	2	G81592	hypothetical prote
14	40	59.7	718	2	T40823	probable para-amin
15	40	59.7	1016	2	T30553	disease resistance
16	40	59.7	1112	2	T10504	disease resistance
17	40	59.7	1148	2	D82091	exodeoxyribonuclea
18	39	58.2	74	2	D69353	hypothetical prote
19	39	58.2	228	2	A29347	alpha-amylase (EC
20	39	58.2	250	2	AD1956	phosphoesterase-re
21	39	58.2	272	2	S52012	cytochrome-c oxida
22	39	58.2	500	2	S72506	lycopene beta-cycl
23	39	58.2	500	2	S66349	lycopene beta-cycl
24	39	58.2	562	2	C72278	hypothetical prote
25	39	58.2	803	2	A86655	hypothetical prote
26	39	58.2	838	2	A96557	probable receptor
27	39	58.2	932	2	T48489	receptor-like prot
28	39	58.2	1049	1	S19421	ATP-dependent perm
29	39	58.2	1121	2	G64103	exodeoxyribonuclea
30	39	58.2	1313	2	T29027	hypothetical prote
31	39	58.2	1506	2	T32909	hypothetical prote
32	38	56.7	169	2	A83134	conserved hypothet
33	38	56.7	191	2	T32278	hypothetical prote
34	38	56.7	194	2	S75835	hypothetical prote

35	38	56.7	222	1	BFBO	folate-binding pro
36	38	56.7	289	2	B90319	hypothetical prote
37	38	56.7	334	2	T23027	hypothetical prote
38	38	56.7	398	2	I53340	paired box transcr
39	38	56.7	667	2	B96575	hypothetical prote
40	38	56.7	669	2	C96814	hypothetical prote
41	38	56.7	675	2	PQ0227	adenylate cyclase
42	38	56.7	1015	2	T15830	hypothetical prote
43	38	56.7	1248	2	A53588	adenylate cyclase
44	38	56.7	1251	2	S48687	type VIII adenylyl
45	37.5	56.0	170	2	B97811	proline/betaine tr

ALIGNMENTS

RESULT 1

T16677

hypothetical protein R04A9.3 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 04-Mar-2000

C;Accession: T16677

R;Geisel, C.

submitted to the EMBL Data Library, November 1995

A;Description: The sequence of *C. elegans* cosmid R04A9.

A;Reference number: Z18558

A;Accession: T16677

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-267 <GEI>

A;Cross-references: EMBL:U41550; NID:g1118045; PID:g1118047; PIDN:AAA83285.1;

CESP:R04A9.3

C;Genetics:

A;Gene: CESP:R04A9.3

A;Introns: 15/3; 44/3; 80/2; 136/3; 160/1; 197/1; 250/1

C;Superfamily: *Caenorhabditis elegans* hypothetical protein R04A9.3

Query Match 65.7%; Score 44; DB 2; Length 267;

Best Local Similarity 62.5%; Pred. No. 14;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CENWWGDV 8

|: ||||:

Db 19 CQAWWGD L 26

RESULT 2

S73550

DNA polymerase III gamma-tau chain dnaX - *Mycoplasma pneumoniae* (strain ATCC 29342)

N;Alternate names: hypothetical protein C12_orf681

C;Species: *Mycoplasma pneumoniae*

A;Variety: ATCC 29342

C;Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999

C;Accession: S73550

R;Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996

A;Title: Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*.
 A;Reference number: S73327; MUID:97105885; PMID:8948633
 A;Accession: S73550
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-681 <HIM>
 A;Cross-references: EMBL:AE000022; GB:U00089; NID:g1673882; PIDN:AAB95872.1; PID:g1673890
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
 C;Genetics:
 A;Gene: dnaX
 A;Genetic code: SGC3
 C;Superfamily: DNA-directed DNA polymerase III gamma chain

Query Match 65.7%; Score 44; DB 2; Length 681;
 Best Local Similarity 77.8%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CENWWGDVC 9
 | | | | |
 Db 63 CLNWNGDVC 71

RESULT 3

AF2481

hypothetical protein all7030 [imported] - *Nostoc* sp. (strain PCC 7120) plasmid pCC7120alpha

C;Species: *Nostoc* sp. PCC 7120

A;Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C;Accession: AF2481

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, M.; Ishikawa, A.; Kawashima, K.; Kimura, T.; Kishida, Y.; Kohara, M.; Matsumoto, M.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing *Cyanobacterium Anabaena* sp. strain PCC 7120.

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AF2481

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1132 <KUR>

A;Cross-references: GB:BA000020; PIDN:BAB78114.1; PID:g17135568; GSPDB:GN00180

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: all7030

A;Genome: plasmid

Query Match 65.7%; Score 44; DB 2; Length 1132;
 Best Local Similarity 62.5%; Pred. No. 51;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CENWWGDV 8
 |::||| |

Db 1065 CDSWWGQV 1072

RESULT 4

B38919

hypothetical protein 2 - human (fragment)

C;Species: Homo sapiens (man)

C;Date: 05-Aug-1994 #sequence_revision 05-Aug-1994 #text_change 04-Mar-2000

C;Accession: B38919

R;Nomura, N.; Miyajima, N.; Kawarabayashi, Y.; Tabata, S.

submitted to the EMBL Data Library, May 1994

A;Description: Prediction of new human genes by entire sequencing of randomly sampled cDNA clones.

A;Reference number: A38919

A;Accession: B38919

A;Molecule type: mRNA

A;Residues: 1-1054 <NOM>

A;Cross-references: EMBL:D25215

C;Superfamily: ubiquitin-protein ligase homology

F;731-1049/Domain: ubiquitin-protein ligase homology <UBI>

Query Match 64.2%; Score 43; DB 2; Length 1054;
Best Local Similarity 62.5%; Pred. No. 67;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ENWWGDVC 9
: ||| ||
Db 549 DNWWSQVC 556

RESULT 5

A45724

pectate lyase (EC 4.2.2.2) - fungus (Fusarium solani)

C;Species: Fusarium solani

C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000

C;Accession: A45724

R;Gonzalez-Candelas, L.; Kolattukudy, P.E.

J. Bacteriol. 174, 6343-6349, 1992

A;Title: Isolation and analysis of a novel inducible pectate lyase gene from the phytopathogenic fungus Fusarium solani f. sp. pisi (Nectria haematococca, mating population VI).

A;Reference number: A45724; MUID:93015682; PMID:1400187

A;Accession: A45724

A;Status: preliminary

A;Molecule type: DNA; protein

A;Residues: 1-242 <GON>

A;Cross-references: GB:M94691; NID:g168155; PIDN:AAA33338.1; PID:g168156

A;Experimental source: isolate T8

A;Note: sequence extracted from NCBI backbone (NCBIN:115473, NCBIP:115474)

C;Keywords: carbon-oxygen lyase

Query Match 61.2%; Score 41; DB 2; Length 242;
Best Local Similarity 83.3%; Pred. No. 35;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 WWGDVC 9
|| |||

Db 102 WWADVC 107

RESULT 6

T24191

hypothetical protein R11D1.10 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C;Accession: T24191

R;Steward, C.

submitted to the EMBL Data Library, June 1996

A;Reference number: Z19850

A;Accession: T24191

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-443 <WIL>

A;Cross-references: EMBL:Z75547; PIDN:CAA99905.2; GSPDB:GN00023; CESP:R11D1.10

A;Experimental source: clone R11D1

C;Genetics:

A;Gene: CESP:R11D1.10

A;Map position: 5

A;Introns: 27/3; 76/3; 135/3; 188/1; 228/3; 318/2; 348/1

Query Match 61.2%; Score 41; DB 2; Length 443;

Best Local Similarity 83.3%; Pred. No. 61;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ENWWGD 7

:|||||

Db 91 KNWWGD 96

RESULT 7

A53506

folate receptor type gamma - human

N;Contains: folate receptor type gamma'

C;Species: *Homo sapiens* (man)

C;Date: 06-Oct-1994 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000

C;Accession: A53506; B53506

R;Shen, F.; Ross, J.F.; Wang, X.; Ratnam, M.

Biochemistry 33, 1209-1215, 1994

A;Title: Identification of a novel folate receptor, a truncated receptor, and receptor type beta in hematopoietic cells: cDNA cloning, expression, immunoreactivity, and tissue specificity.

A;Reference number: A53506; MUID:94153905; PMID:8110752

A;Accession: A53506

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-243 <SHE>

A;Cross-references: GB:Z32564; NID:g473235; PIDN:CAA83553.1; PID:g473236

A;Experimental source: CML patient, spleen and bone marrow

A;Note: sequence extracted from NCBI backbone (NCBIN:145218, NCBIP:145219)

A;Accession: B53506

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-104 <SH2>

A;Cross-references: GB:Z32633; NID:g474060; PIDN:CAA83566.1; PID:g474061

A;Experimental source: CML patient, spleen and bone marrow
A;Note: sequence extracted from NCBI backbone (NCBIN:145220, NCBIP:145221)
C;Genetics:
A;Gene: GDB:FOLR3
A;Cross-references: GDB:306562
C;Superfamily: folate-binding protein

Query Match 59.7%; Score 40; DB 2; Length 243;
Best Local Similarity 71.4%; Pred. No. 50;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CENWWGD 7
|| || |
Db 137 CERWWED 143

RESULT 8

A45753

folate-binding protein precursor - human

N;Contains: folate receptor; tumor-associated antigen

C;Species: Homo sapiens (man)

C;Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 13-Aug-1999

C;Accession: A44904; B44904; A36515; A45753; S21763; S24405; S47554; A32864;
A47570; A28316; B28316

R;Coney, L.R.; Tomassetti, A.; Carayannopoulos, L.; Frasca, V.; Kamen, B.A.;
Colnaghi, M.I.; Zurawski Jr., V.R.

Cancer Res. 51, 6125-6132, 1991

A;Title: Cloning of a tumor-associated antigen: MOv18 and MOv19 antibodies
recognize a folate-binding protein.

A;Reference number: A44904; MUID:92034730; PMID:1840502

A;Accession: A44904

A;Molecule type: mRNA

A;Residues: 1-257 <CON>

A;Cross-references: GB:U20391; NID:g1483626; PIDN:AAB05827.1; PID:g1483627

A;Experimental source: ovarian carcinoma cell line IGROV1

A;Note: sequence extracted from NCBI backbone (NCBIN:66569, NCBIP:66571)

A;Accession: B44904

A;Molecule type: protein

A;Residues: 31-36, 'X', 38-41, 'WX', 44-47, 'X', 49, 'X', 51-53, 'XX', 56, 'X' <CO2>

A;Note: sequence extracted from NCBI backbone (NCBIP:66567)

R;Elwood, P.C.

J. Biol. Chem. 264, 14893-14901, 1989

A;Title: Molecular cloning and characterization of the human folate-binding
protein cDNA from placenta and malignant tissue culture (KB) cells.

A;Reference number: A36515; MUID:89359294; PMID:2768245

A;Accession: A36515

A;Molecule type: mRNA

A;Residues: 1-45, 'R', 47-257 <ELW>

A;Cross-references: GB:J05013

A;Experimental source: nasopharyngeal epidermoid carcinoma cell line KB

A;Note: the authors translated the codon AGG for residue 46 as Lys

R;Lacey, S.W.; Sanders, J.M.; Rothberg, K.G.; Anderson, R.G.W.; Kamen, B.A.

J. Clin. Invest. 84, 715-720, 1989

A;Title: Complementary DNA for the folate binding protein correctly predicts
anchoring to the membrane by glycosyl-phosphatidylinositol.

A;Reference number: A45753; MUID:89340896; PMID:2527252

A;Accession: A45753

A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-257 <LAC>
 A;Cross-references: GB:M28099; NID:g182415; PIDN:AAA35822.1; PID:g182416
 R;Sadasivan, E.; Cedeno, M.; Rothenberg, S.P.
 Biochim. Biophys. Acta 1131, 91-94, 1992
 A;Title: Genomic organization of the gene and a related pseudogene for a human folate binding protein.
 A;Reference number: S21763; MUID:92256496; PMID:1581364
 A;Accession: S21763
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-257 <SAD>
 R;Campbell, I.G.; Jones, T.A.; Foulkes, W.D.; Trowsdale, J.
 Cancer Res. 51, 5329-5338, 1991
 A;Title: Folate-binding protein is a marker for ovarian cancer.
 A;Reference number: S24405; MUID:92005454; PMID:1717147
 A;Accession: S24405
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-257 <CAM>
 A;Cross-references: EMBL:X62753; NID:g28428; PIDN:CAA44610.1; PID:g28429
 R;Prasad, P.D.; Ramamoorthy, S.; Moe, A.J.; Smith, C.H.; Leibach, F.H.; Ganapathy, V.
 Biochim. Biophys. Acta 1223, 71-75, 1994
 A;Title: Selective expression of the high-affinity isoform of the folate receptor (FR-alpha) in the human placental syncytiotrophoblast and choriocarcinoma cells.
 A;Reference number: S47554; MUID:94339186; PMID:8061055
 A;Accession: S47554
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-105,162-257 <PRA>
 R;Sadasivan, E.; Rothenberg, S.P.
 J. Biol. Chem. 264, 5806-5811, 1989
 A;Title: The complete amino acid sequence of a human folate binding protein from KB cells determined from the cDNA.
 A;Reference number: A32864; MUID:89174638; PMID:2538429
 A;Accession: A32864
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 24-183,'S',185-249 <SA3>
 A;Cross-references: GB:M25317; NID:g182421; PIDN:AAA74896.1; PID:g182422
 R;Sadasivan, E.; Rothenberg, S.P.
 Proc. Soc. Exp. Biol. Med. 189, 240-244, 1988
 A;Title: Molecular cloning of the complementary DNA for a human folate binding protein (42804).
 A;Reference number: A47570; MUID:89057954; PMID:3194438
 A;Accession: A47570
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 24-45 <SA2>
 A;Cross-references: EMBL:M35069; NID:g182419; PIDN:AAA35824.1; PID:g182420
 R;Luhrs, C.A.; Pitiranggon, P.; da Costa, M.; Rothenberg, S.P.; Slomiany, B.L.; Brink, L.; Tous, G.I.; Stein, S.
 Proc. Natl. Acad. Sci. U.S.A. 84, 6546-6549, 1987

A;Title: Purified membrane and soluble folate binding proteins from cultured KB cells have similar amino acid compositions and molecular weights but differ in fatty acid acylation.

A;Reference number: A28316; MUID:87317689; PMID:3476960

A;Accession: A28316

A;Status: preliminary

A;Molecule type: protein

A;Residues: 26-36,'X',38-43 <LUH>

A;Experimental source: KB cells

C;Genetics:

A;Gene: GDB:FOLR1; FOLR

A;Cross-references: GDB:128061; OMIM:136430

A;Map position: 11q13.3-11q14.1

C;Superfamily: folate-binding protein

F;1-25/Domain: signal sequence #status predicted <SIG>

F;26-257/Product: folate-binding protein #status predicted <MAT>

F;31-257/Product: tumor-associated antigen #status experimental <ANT>

Query Match 59.7%; Score 40; DB 2; Length 257;
Best Local Similarity 71.4%; Pred. No. 53;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CENWWGD 7
|| || |
Db 139 CEQWWED 145

RESULT 9

C64567

fucosyltransferase - Helicobacter pylori (strain 26695)

C;Species: Helicobacter pylori

C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999

C;Accession: C64567

R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk, H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.; Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997

A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.M.; Venter, J.C.

A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A;Reference number: A64520; MUID:97394467; PMID:9252185

A;Accession: C64567

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-425 <TOM>

A;Cross-references: GB:AE000554; GB:AE000511; NID:g2313475; PIDN:AAD07447.1;

PID:g2313482; TIGR:HP0379

Query Match 59.7%; Score 40; DB 2; Length 425;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 NWWGD 7
 |||||
Db 33 NWWGD 37

RESULT 10

C64601

fucosyltransferase - *Helicobacter pylori* (strain 26695)

C;Species: *Helicobacter pylori*

C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999

C;Accession: C64601

R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.;
Fleischmann, R.D.; Ketchum, K.A.; Klenk, H.P.; Gill, S.; Dougherty, B.A.;
Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, K.; Fitzgerald,
L.M.; Lee, N.; Adams, M.D.; Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback,
T.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.;
Bowman, C.; Watthey, L.

Nature 388, 539-547, 1997

A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.;
Fraser, C.M.; Venter, J.C.

A;Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.

A;Reference number: A64520; MUID:97394467; PMID:9252185

A;Accession: C64601

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-476 <TOM>

A;Cross-references: GB:AE000578; GB:AE000511; NID:g2313759; PIDN:AAD07710.1;
PID:g2313769; TIGR:HP0651

Query Match 59.7%; Score 40; DB 2; Length 476;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 NWWGD 7
 |||||
Db 33 NWWGD 37

RESULT 11

D72075

hypothetical protein - *Chlamydophila pneumoniae* (strain CWL029)

C;Species: *Chlamydophila pneumoniae*, *Chlamydia pneumoniae*

C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 02-Sep-2000

C;Accession: D72075

R;Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.;
Grimwood, J.; Davis, R.W.; Stephens, R.S.

Nature Genet. 21, 385-389, 1999

A;Title: Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*.

A;Reference number: A72000; MUID:99206606; PMID:10192388

A;Accession: D72075

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-582 <ARN>

A;Cross-references: GB:AE001630; GB:AE001363; NID:g4376740; PIDN:AAD18599.1;
PID:g4376741

A;Experimental source: strain CWL029
C;Genetics:
A;Gene: CPn0457
C;Superfamily: Chlamydia hypothetical protein CPn0462

Query Match 59.7%; Score 40; DB 2; Length 582;
Best Local Similarity 66.7%; Pred. No. 1.1e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CENWWG 6
|::|||
Db 426 CDSWWG 431

RESULT 12

E86547
hypothetical protein CPj0457 [imported] - Chlamydophila pneumoniae (strain J138)
C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C;Accession: E86547
R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.;
Shiba, T.; Ishii, K.; Hattori, M.; Kuhara, S.; Nakazawa, T.
Nucleic Acids Res. 28, 2311-2314, 2000
A;Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A;Reference number: A86491; MUID:20330349; PMID:10871362
A;Accession: E86547
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-629 <STO>
A;Cross-references: GB:BA000008; NID:g8978827; PIDN:BAA98663.1; GSPDB:GN00142
A;Experimental source: strain J138
C;Genetics:
A;Gene: CPj0457
C;Superfamily: Chlamydia hypothetical protein CPn0462

Query Match 59.7%; Score 40; DB 2; Length 629;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CENWWG 6
|::|||
Db 426 CDSWWG 431

RESULT 13

G81592
hypothetical protein CP0295 [imported] - Chlamydophila pneumoniae (strain AR39)
C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Sep-2000
C;Accession: G81592
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.;
Hickey, E.K.; Peterson, J.; Utterback, T.; Berry, K.; Bass, S.; Linher, K.;
Weidman, J.; Khouri, H.; Craven, B.; Bowman, C.; Dodson, R.; Gwinn, M.; Nelson,
W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, S.L.; Eisen, J.; Fraser,
C.M.
Nucleic Acids Res. 28, 1397-1406, 2000

A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.

A;Reference number: A81500; MUID:20150255; PMID:10684935

A;Accession: G81592

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-629 <REA>

A;Cross-references: GB:AE002191; GB:AE002161; NID:g7189216; PIDN:AAF38152.1;

PID:g7189221; GSPDB:GN00122; TIGR:CP0295

A;Experimental source: strain AR39, HL cells

C;Genetics:

A;Gene: CP0295

C;Superfamily: Chlamydia hypothetical protein CPn0462

Query Match 59.7%; Score 40; DB 2; Length 629;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CENWWG 6
|::|||
Db 426 CDSWWG 431

RESULT 14

T40823

probable para-aminobenzoate synthase - fission yeast (Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Jan-2000

C;Accession: T40823

R;Beck, A.; Reinhardt, R.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, October 1998

A;Reference number: Z21949

A;Accession: T40823

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-718 <BEC>

A;Cross-references: EMBL:AL032684; PIDN:CAA21814.1; GSPDB:GN00067;

SPDB:SPBP8B7.29

A;Experimental source: strain 972h-; clone p1 p8B7

C;Genetics:

A;Gene: SPDB:SPBP8B7.29

A;Map position: 2

C;Superfamily: yeast p-aminobenzoate synthase; trpG homology

Query Match 59.7%; Score 40; DB 2; Length 718;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CENWWGDV 8
| |||::
Db 404 CSEWWGEL 411

RESULT 15

T30553

disease resistance protein Hcr2-5D - tomato

C;Species: Lycopersicon esculentum (tomato)

C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
C;Accession: T30553
R;Dixon, M.S.; Hatzixanthis, K.; Jones, D.A.; Harrison, K.; Jones, J.D.G.
Plant Cell 10, 1915-1926, 1998
A;Title: The tomato Cf-5 disease resistance gene and six homologues show
pronounced allelic variation in leucine-rich repeat copy number.
A;Reference number: Z20855; MUID:99030197; PMID:9811798
A;Accession: T30553
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1016 <DIX>
A;Cross-references: EMBL:AF053998; NID:g3894392; PID:g3894393; PIDN:AAC78596.1
C;Genetics:
A;Note: Hcr2-5D

Query Match 59.7%; Score 40; DB 2; Length 1016;
Best Local Similarity 55.6%; Pred. No. 1.8e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CENWWGDVC 9
|::|:| ||
Db 59 CKDWYGVVC 67

Search completed: November 13, 2003, 09:52:50
Job time : 10.375 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 5.15625 Seconds
(without alignments)
82.083 Million cell updates/sec

Title: US-09-228-866-2
Perfect score: 67
Sequence: 1 CENWWGDVC 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	65.7	681	1	DP3X_MYCPN	P75177 mycoplasma
2	43	64.2	1050	1	HER3_HUMAN	Q15034 homo sapien
3	42	62.7	1151	1	XPO4_HUMAN	Q9c0e2 homo sapien
4	42	62.7	1151	1	XPO4_MOUSE	Q9esj0 mus musculu
5	40	59.7	243	1	FOL3_HUMAN	P41439 homo sapien
6	40	59.7	257	1	FOL1_HUMAN	P15328 homo sapien
7	40	59.7	455	1	ANL3_MOUSE	Q9r182 mus musculu
8	39	58.2	489	1	AMY_TRICA	P09107 tribolium c
9	39	58.2	500	1	LCYB_TOBAC	Q43578 nicotiana t
10	39	58.2	853	1	ENV_HV1EL	P04581 human immun
11	39	58.2	1049	1	ADP1_YEAST	P25371 saccharomyc
12	39	58.2	1121	1	EX5C_HAEIN	P44945 haemophilus
13	38	56.7	222	1	FOL1_BOVIN	P02702 bos taurus
14	38	56.7	1248	1	CYA8_RAT	P40146 rattus norv
15	38	56.7	1249	1	CYA8_MOUSE	P97490 mus musculu
16	38	56.7	1251	1	CYA8_HUMAN	P40145 homo sapien
17	37	55.2	180	1	XG_HUMAN	P55808 homo sapien
18	37	55.2	255	1	FOL1_MOUSE	P35846 mus musculu
19	37	55.2	255	1	FOL2_HUMAN	P14207 homo sapien
20	37	55.2	296	1	STC2_MOUSE	O88452 mus musculu
21	37	55.2	296	1	STC2_RAT	Q9r0k8 rattus norv
22	37	55.2	302	1	STC2_HUMAN	O76061 homo sapien
23	37	55.2	302	1	STC2_MACNE	O97561 macaca neme
24	37	55.2	349	1	YGBO_ECOLI	Q57261 escherichia
25	37	55.2	610	1	VE1_HPV60	Q80943 human papil
26	37	55.2	643	1	UROM_BOVIN	P48733 bos taurus
27	37	55.2	1111	1	MATE_MOUSE	Q9r1m5 mus musculu
28	37	55.2	1826	1	SUIS_HUMAN	P14410 homo sapien
29	37	55.2	3014	1	CLR1_HUMAN	Q9nyq6 homo sapien
30	37	55.2	3034	1	CLR1_MOUSE	O35161 mus musculu
31	36.5	54.5	616	1	TR11_HUMAN	Q9y6q6 homo sapien
32	36.5	54.5	636	1	S6A7_HUMAN	Q99884 homo sapien
33	36.5	54.5	833	1	SRC2_MOUSE	P59222 mus musculu
34	36.5	54.5	870	1	SRC2_HUMAN	Q96gp6 homo sapien
35	36.5	54.5	1408	1	SERR_DROME	P18168 drosophila
36	36	53.7	46	1	AX1_BETVU	P81493 beta vulgar
37	36	53.7	197	1	STC2_CAVPO	P57675 cavia porce
38	36	53.7	260	1	COX3_ASTPE	Q33824 asterina pe
39	36	53.7	376	1	CPR2_ARATH	Q9lxw3 arabidopsis
40	36	53.7	415	1	FTSW_MESVI	Q9mum4 mesostigma
41	36	53.7	417	1	TNAB_PROVU	P28785 proteus vul
42	36	53.7	465	1	LIPP_CAVPO	P50903 cavia porce
43	36	53.7	465	1	LIPP_RAT	P27657 rattus norv
44	36	53.7	539	1	TYRO_ASPOR	Q00234 aspergillus
45	36	53.7	605	1	SYA_TREPA	O83980 treponema p

ALIGNMENTS

RESULT 1

DP3X_MYCPN

ID DP3X_MYCPN STANDARD; PRT; 681 AA.

AC P75177;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE DNA polymerase III subunit gamma/tau (EC 2.7.7.7).

GN DNAX OR MPN618 OR MP224.

OS Mycoplasma pneumoniae.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

OX NCBI_TaxID=2104;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 29342 / M129;

RX MEDLINE=97105885; PubMed=8948633;

RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,

RA Herrmann R.;

RT "Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae.";

RL Nucleic Acids Res. 24:4420-4449(1996).

CC -!- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.

CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.

CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
+ {DNA}(N).

CC -!- SUBUNIT: DNA polymerase III contains a core (composed of alpha,
epsilon and theta chains) that associates with a tau subunit. This
core dimerizes to form the POLIII' complex. PolIII' associates
with the gamma complex (composed of gamma, delta, delta', psi and
chi chains) and with the beta chain to form the complete DNA
polymerase III complex (By similarity).

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AE000022; AAB95872.1; -.

DR PIR; S73550; S73550.

DR InterPro; IPR003593; AAA_ATPase.

DR InterPro; IPR003959; AAA_ATPase_cent.

DR InterPro; IPR000862; RFCdomain.

DR Pfam; PF00004; AAA; 1.

DR SMART; SM00382; AAA; 1.

KW Transferase; DNA-directed DNA polymerase; DNA replication;

KW ATP-binding; Complete proteome.

FT NP_BIND 44 51 ATP (POTENTIAL).

SQ SEQUENCE 681 AA; 76212 MW; E3DDC6A580FFCBCC CRC64;

Query Match 65.7%; Score 44; DB 1; Length 681;
Best Local Similarity 77.8%; Pred. No. 11;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CENWWGDVC 9
| || |||
Db 63 CLNWNVDVC 71

RESULT 2

HER3_HUMAN

ID HER3_HUMAN STANDARD; PRT; 1050 AA.
AC Q15034;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE HECT domain and RCC1-like domain protein 3.
GN HERC3 OR KIAA0032.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96051387; PubMed=7584026;
RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayashi Y.,
RA Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. I.
RT The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
RT analysis of randomly sampled cDNA clones from human immature myeloid
RT cell line KG-1.";
RL DNA Res. 1:27-35(1994).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=21099818; PubMed=11163799;
RA Cruz C., Ventura F., Bartrons R., Rosa J.L.;
RT "HERC3 binding to and regulation by ubiquitin.";
RL FEBS Lett. 488:74-80(2001).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic. Also found in vesicular-like
CC structures.
CC -!- PTM: Substrate of ubiquitination and is degraded by the
CC proteasome.
CC -!- SIMILARITY: Contains 1 HECT-type E3 ubiquitin-protein ligase
CC domain.
CC -!- SIMILARITY: Contains 7 RCC1 repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D25215; BAA04945.1; -.
DR Genew; HGNC:4876; HERC3.
DR MIM; 605200; -.
DR InterPro; IPR000569; HECT_domain.

DR InterPro; IPR000408; Reg_chr_condens.
 DR Pfam; PF00632; HECT; 1.
 DR Pfam; PF00415; RCC1; 4.
 DR PRINTS; PR00633; RCCNDNSATION.
 DR SMART; SM00119; HECTc; 1.
 DR PROSITE; PS50237; HECT; 1.
 DR PROSITE; PS00625; RCC1_1; FALSE_NEG.
 DR PROSITE; PS00626; RCC1_2; 4.
 DR PROSITE; PS50012; RCC1_3; 7.
 KW Ubl conjugation pathway; Ubl conjugation; Repeat.
 FT REPEAT 1 51 RCC1 1.
 FT REPEAT 52 101 RCC1 2.
 FT REPEAT 102 154 RCC1 3.
 FT REPEAT 156 207 RCC1 4.
 FT REPEAT 208 259 RCC1 5.
 FT REPEAT 261 311 RCC1 6.
 FT REPEAT 313 366 RCC1 7.
 FT DOMAIN 951 1050 HECT.
 FT BINDING 1018 1018 UBIQUITIN (BY SIMILARITY).
 SQ SEQUENCE 1050 AA; 117188 MW; 5F08A1DE1F40B912 CRC64;

Query Match 64.2%; Score 43; DB 1; Length 1050;
 Best Local Similarity 62.5%; Pred. No. 25;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ENWWGDVC 9
 :||| ||
 Db 545 DNWSQVC 552

RESULT 3

XPO4_HUMAN

ID XPO4_HUMAN STANDARD; PRT; 1151 AA.
 AC Q9C0E2; Q9H934;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Exportin 4 (Exp4).
 GN XPO4 OR KIAA1721.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=21082932; PubMed=11214970;
 RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIX.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 7:347-355(2000).
 RN [2]
 RP SEQUENCE OF 337-1151 FROM N.A.
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,
 RA Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y.,

RA Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K.,
 RA Masuho Y., Kanehori K.;
 RT "NEDO human cDNA sequencing project.";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
 CC -!- FUNCTION: Mediates nuclear export of eIF-5A (eukaryotic
 CC translation initiation factor 5A) and possibly that of other
 CC cargoes (By similarity).
 CC -!- SUBUNIT: Binds to GTP-bound form of Ran (By similarity).
 CC -!- SUBCELLULAR LOCATION: Nuclear; once bound to eIF-5A and Ran
 CC the complex translocates to the cytoplasm (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE EXPORTIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB051508; BAB21812.1; -.
 DR EMBL; AK023108; BAB14409.1; ALT_INIT.
 DR Genew; HGNC:17796; XPO4.
 KW Nuclear protein; Transport; Protein transport.
 FT CONFLICT 511 511 L -> S (IN REF. 2).
 SQ SEQUENCE 1151 AA; 130139 MW; 38E7EEFC938B07C5 CRC64;

Query Match 62.7%; Score 42; DB 1; Length 1151;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CENWW 5
 |||||
 Db 723 CENWW 727

RESULT 4

XPO4_MOUSE
 ID XPO4_MOUSE STANDARD; PRT; 1151 AA.
 AC Q9ESJ0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Exportin 4 (Exp4).
 GN XPO4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=20402342; PubMed=10944119;
 RA Lipowsky G., Bischoff F.R., Schwarzmaier P., Kraft R., Kostka S.,
 RA Hartmann E., Kutay U., Goerlich D.;
 RT "Exportin 4: a mediator of a novel nuclear export pathway in higher
 RT eukaryotes.";
 RL EMBO J. 19:4362-4371(2000).

CC -!- FUNCTION: Mediates nuclear export of eIF-5A (eukaryotic
CC translation initiation factor 5A) and possibly that of other
CC cargoes.
CC -!- SUBUNIT: Binds to GTP-bound form of Ran.
CC -!- SUBCELLULAR LOCATION: Nuclear; once bound to eIF-5A and Ran
CC the complex translocates to the cytoplasm.
CC -!- SIMILARITY: BELONGS TO THE EXPORTIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF145021; AAG09133.1; -.
DR MGD; MGI:1888526; Xpo4.
KW Nuclear protein; Transport; Protein transport.
SQ SEQUENCE 1151 AA; 129964 MW; 5836A4940EB598BE CRC64;

Query Match 62.7%; Score 42; DB 1; Length 1151;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CENWW 5
 |||||
Db 723 CENWW 727

RESULT 5

FOL3_HUMAN

ID FOL3_HUMAN STANDARD; PRT; 243 AA.
AC P41439;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Folate receptor gamma precursor (FR-gamma) (Folate receptor 3).
GN FOLR3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Hematopoietic;
RX MEDLINE=94153905; PubMed=8110752;
RA Shen F., Ross J.F., Wang X., Ratnam M.;
RT "Identification of a novel folate receptor, a truncated receptor, and
RT receptor type beta in hematopoietic cells: cDNA cloning, expression,
RT immunoreactivity, and tissue specificity.";
RL Biochemistry 33:1209-1215(1994).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95244494; PubMed=7727426;
RA Shen F., Wu M., Ross J.F., Miller D., Ratnam M.;
RT "Folate receptor type gamma is primarily a secretory protein due to

RT lack of an efficient signal for glycosylphosphatidylinositol
 RT modification: protein characterization and cell type specificity.";
 RL Biochemistry 34:5660-5665(1995).
 CC -!- FUNCTION: BINDS TO FOLATE AND REDUCED FOLIC ACID DERIVATIVES AND
 CC MEDIATES DELIVERY OF 5-METHYLTETRAHYDROFOLATE TO THE INTERIOR OF
 CC CELLS.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=Long;
 CC IsoId=P41439-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=P41439-2; Sequence=VSP_001506;
 CC -!- TISSUE SPECIFICITY: SPLEEN, THYMUS, BONE MARROW, OVARIAN
 CC CARCINOMA, AND UTERINE CARCINOMA.
 CC -!- PTM: EIGHT DISULFIDE BONDS ARE PRESENT (PROBABLE).
 CC -!- SIMILARITY: BELONGS TO THE FOLATE RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z32564; CAA83553.1; -.
 DR EMBL; Z32633; CAA83566.1; -.
 DR EMBL; U08471; AAA18382.1; -.
 DR EMBL; U08470; AAA18381.1; -.
 DR PIR; A53506; A53506.
 DR Genew; HGNC:3795; FOLR3.
 DR MIM; 602469; -.
 DR GO; GO:0005624; C:membrane fraction; TAS.
 DR GO; GO:0005542; F:folic acid binding activity; TAS.
 DR GO; GO:0015025; F:GPI-anchored membrane-bound receptor; TAS.
 DR GO; GO:0015884; P:folate transport; TAS.
 DR InterPro; IPR004269; Folate_rec.
 DR Pfam; PF03024; Folate_rec; 1.
 KW Receptor; Glycoprotein; Signal; Folate-binding; Multigene family;
 KW Alternative splicing.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 243 FOLATE RECEPTOR GAMMA.
 FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 159 159 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 105 243 Missing (in isoform Short).
 FT /FTid=VSP_001506.
 SQ SEQUENCE 243 AA; 27638 MW; AC7636EB5355647B CRC64;

Query Match 59.7%; Score 40; DB 1; Length 243;
 Best Local Similarity 71.4%; Pred. No. 17;
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CENWWGD 7
 || || |
 Db 137 CERWWED 143

RESULT 6

FOL1_HUMAN

ID FOL1_HUMAN STANDARD; PRT; 257 AA.
AC P15328;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Folate receptor alpha precursor (FR-alpha) (Folate receptor 1) (Folate
DE receptor, adult) (Adult folate-binding protein) (FBP) (Ovarian tumor-
DE associated antigen MOv18) (KB cells FBP).
GN FOLR1 OR FOLR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89340896; PubMed=2527252;
RA Lacey S.W., Sanders J.M., Rothberg K.G., Anderson R.G.W.,
RA Kamen B.A.;
RT "Complementary DNA for the folate binding protein correctly predicts
RT anchoring to the membrane by glycosyl-phosphatidylinositol.";
RL J. Clin. Invest. 84:715-720(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89359294; PubMed=2768245;
RA Elwood P.C.;
RT "Molecular cloning and characterization of the human folate-binding
RT protein cDNA from placenta and malignant tissue culture (KB) cells.";
RL J. Biol. Chem. 264:14893-14901(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=92005454; PubMed=1717147;
RA Campbell I.G., Jones T.A., Foulkes W.D., Trowsdale J.;
RT "Folate-binding protein is a marker for ovarian cancer.";
RL Cancer Res. 51:5329-5338(1991).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=92034730; PubMed=1840502;
RA Coney L.R., Tomassetti A., Carayannopoulos L., Frasca V.,
RA Kamen B.A., Colnaghi M.I., Zurawski V.R. Jr.;
RT "Cloning of a tumor-associated antigen: MOv18 and MOv19 antibodies
RT recognize a folate-binding protein.";
RL Cancer Res. 51:6125-6132(1991).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=92256496; PubMed=1581364;
RA Sadasivan E., Cedeno M., Rothenberg S.P.;
RT "Genomic organization of the gene and a related pseudogene for a
RT human folate binding protein.";
RL Biochim. Biophys. Acta 1131:91-94(1992).
RN [6]
RP SEQUENCE OF 24-249 FROM N.A.

RX MEDLINE=89174638; PubMed=2538429;
 RA Sadasivan E., Rothenberg S.P.;
 RT "The complete amino acid sequence of a human folate binding protein
 RT from KB cells determined from the cDNA.";
 RL J. Biol. Chem. 264:5806-5811(1989).
 RN [7]
 RP SEQUENCE OF 26-43.
 RX MEDLINE=87317689; PubMed=3476960;
 RA Luhrs C.A., Pitiranggon P., da Costa M., Rothenberg S.P.,
 RA Slomiany B.L., Brink L., Tous G.I., Stein S.;
 RT "Purified membrane and soluble folate binding proteins from cultured
 RT KB cells have similar amino acid compositions and molecular weights
 RT but differ in fatty acid acylation.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:6546-6549(1987).
 RN [8]
 RP GPI-ANCHOR.
 RX MEDLINE=96062525; PubMed=7578066;
 RA Yan W., Ratnam M.;
 RT "Preferred sites of glycosylphosphatidylinositol modification in
 RT folate receptors and constraints in the primary structure of the
 RT hydrophobic portion of the signal.";
 RL Biochemistry 34:14594-14600(1995).
 CC -!- FUNCTION: BINDS TO FOLATE AND REDUCED FOLIC ACID DERIVATIVES AND
 CC MEDIATES DELIVERY OF 5-METHYLTETRAHYDROFOLATE TO THE INTERIOR OF
 CC CELLS.
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1; Synonyms=GPI-anchored;
 CC IsoId=P15328-1; Sequence=Displayed;
 CC Name=2; Synonyms=Cytoplasmic;
 CC IsoId=P15328-2; Sequence=Not described;
 CC -!- TISSUE SPECIFICITY: FR-ALPHA LEVELS ARE GREATLY ELEVATED IN A
 CC VARIETY OF MALIGNANT TISSUES OF EPITHELIAL ORIGIN COMPARED WITH
 CC NORMAL TISSUES.
 CC -!- PTM: EIGHT DISULFIDE BONDS ARE PRESENT (PROBABLE).
 CC -!- SIMILARITY: BELONGS TO THE FOLATE RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M28099; AAA35822.1; -.
 DR EMBL; X62753; CAA44610.1; -.
 DR EMBL; J05013; AAA35823.1; -.
 DR EMBL; M25317; AAA74896.1; -.
 DR EMBL; U20391; AAB05827.1; -.
 DR PIR; A44904; A45753.
 DR Genew; HGNC:3791; FOLR1.
 DR MIM; 136430; -.
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.
 DR GO; GO:0005624; C:membrane fraction; TAS.
 DR GO; GO:0005542; F:folic acid binding activity; TAS.

DR GO; GO:0015025; F:GPI-anchored membrane-bound receptor; TAS.
 DR GO; GO:0015884; P:folate transport; TAS.
 DR GO; GO:0006898; P:receptor mediated endocytosis; TAS.
 DR InterPro; IPR004269; Folate_rec.
 DR Pfam; PF03024; Folate_rec; 1.
 KW Receptor; Glycoprotein; Signal; Folate-binding; Membrane; GPI-anchor;
 KW Alternative splicing; Polymorphism.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 234 FOLATE RECEPTOR ALPHA.
 FT PROPEP 235 257 REMOVED IN MATURE FORM.
 FT LIPID 234 234 GPI-ANCHOR.
 FT CARBOHYD 69 69 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
 FT CARBOHYD 161 161 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
 FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 160 160 W -> C (IN dbSNP:1801932).
 FT /FTId=VAR_011963.
 FT CONFLICT 184 184 T -> S (IN REF. 6).
 SQ SEQUENCE 257 AA; 29819 MW; D458D8BB047C96A6 CRC64;

Query Match 59.7%; Score 40; DB 1; Length 257;
 Best Local Similarity 71.4%; Pred. No. 18;
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CENWWGD 7
 || || |
 Db 139 CEQWWED 145

RESULT 7

ANL3_MOUSE

ID ANL3_MOUSE STANDARD; PRT; 455 AA.
 AC Q9R182;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Angiopoietin-related protein 3 precursor (Angiopoietin-like 3).
 GN ANGPTL3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20112762; PubMed=10644446;
 RA Conklin D., Gilbertson D., Taft D.W., Maurer M.F., Whitmore T.E.,
 RA Smith D.L., Walker K.M., Chen L.H., Wattler S., Nehls M., Lewis K.B.;
 RT "Identification of a mammalian angiopoietin-related protein expressed
 RT specifically in liver."
 RL Genomics 62:477-482(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
 CC -!- SIMILARITY: Contains 1 fibrinogen C-terminal domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF162224; AAD45920.1; -.
 DR EMBL; BC019491; AAH19491.1; -.
 DR HSSP; P02671; 1FZD.
 DR MGD; MGI:1353627; Angptl3.
 DR InterPro; IPR002181; Fibrinogen_C.
 DR Pfam; PF00147; fibrinogen_C; 1.
 DR SMART; SM00186; FBG; 1.
 DR PROSITE; PS00514; FIBRIN_AG_C DOMAIN; FALSE_NEG.
 KW Signal; Coiled coil; Glycoprotein.
 FT SIGNAL 1 16 POTENTIAL.
 FT CHAIN 17 455 ANGIOPOIETIN-RELATED PROTEIN 3.
 FT DOMAIN 85 206 COILED COIL (POTENTIAL).
 FT DOMAIN 241 455 FIBRINOGEN C-TERMINAL.
 FT DISULFID 246 274 BY SIMILARITY.
 FT DISULFID 394 408 BY SIMILARITY.
 FT CARBOHYD 115 115 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 232 232 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 357 357 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 455 AA; 52543 MW; 31609D3700D3F33D CRC64;

Query Match 59.7%; Score 40; DB 1; Length 455;
 Best Local Similarity 66.7%; Pred. No. 32;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 4 WWGDVC 9
 |||:|
 Db 403 WWNDIC 408

RESULT 8

AMY_TRICA

ID AMY_TRICA STANDARD; PRT; 489 AA.

AC P09107;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Alpha-amylase precursor (EC 3.2.1.1) (1,4-alpha-D-glucan

DE glucanohydrolase) (Fragment).

OS Tribolium castaneum (Red flour beetle).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;

OC Tenebrionidae; Tribolium.

OX NCBI_TaxID=7070;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=88200288; PubMed=3129570;

RA Hickey D.A., Benkel B.F., Boer P.H., Genest Y., Abukashawa S.,

RA Ben-David G.;

RT "Enzyme-coding genes as molecular clocks: the molecular evolution of

RT animal alpha-amylases.";

RL J. Mol. Evol. 26:252-256(1987).

CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic

CC linkages in oligosaccharides and polysaccharides.

CC -!- COFACTOR: BINDS A CALCIUM ION REQUIRED FOR ITS ACTIVITY.

CC -!- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO

CC KNOWN AS THE ALPHA-AMYLASE FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X06905; CAA30009.1; -.

DR PIR; A29347; A29347.

DR HSSP; P56634; 1JAE.

DR InterPro; IPR006589; Alp_amyl_cat_sub.

DR InterPro; IPR006048; Alpha_amyl_C.

DR InterPro; IPR006047; Alpha_amyl_cat.

DR InterPro; IPR006046; Glyco_hydro_13.

DR Pfam; PF00128; alpha-amylase; 1.

DR Pfam; PF02806; alpha-amylase_C; 1.

DR PRINTS; PR00110; ALPHAAMYLASE.

DR SMART; SM00642; Aamy; 1.

DR SMART; SM00632; Aamy_C; 1.

KW Hydrolase; Glycosidase; Carbohydrate metabolism; Calcium; Signal.

FT NON_TER 1 1

FT SIGNAL <1 16 POTENTIAL.

FT CHAIN 17 489 ALPHA-AMYLASE.

FT ACT_SITE 203 203 BY SIMILARITY.

FT ACT_SITE 207 207 BY SIMILARITY.

FT ACT_SITE 305 305 BY SIMILARITY.

FT DISULFID 44 102 BY SIMILARITY.

FT DISULFID 152 166 BY SIMILARITY.

FT DISULFID 372 378 BY SIMILARITY.
FT DISULFID 443 455 BY SIMILARITY.
SQ SEQUENCE 489 AA; 53247 MW; D1AB107C48FF8721 CRC64;

Query Match 58.2%; Score 39; DB 1; Length 489;
Best Local Similarity 83.3%; Pred. No. 49;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ENWWGD 7
|||
Db 401 ENWWS D 406

RESULT 9

LCYB_TOBAC

ID LCYB_TOBAC STANDARD; PRT; 500 AA.
AC Q43578;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lycopene beta cyclase, chloroplast precursor (EC 1.14.--).
GN LCY1 OR CRTL-1.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiids; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Samsun NN; TISSUE=Leaf;
RX MEDLINE=96434545; PubMed=8837512;
RA Cunningham F.X. Jr., Pogson B., Sun Z., McDonald K.A., Dellapenna D.,
RA Gantt E.;
RT "Functional analysis of the beta and epsilon lycopene cyclase enzymes
RT of Arabidopsis reveals a mechanism for control of cyclic carotenoid
RT formation.";
RL Plant Cell 8:1613-1626(1996).
CC -!- FUNCTION: CATALYZES THE DOUBLE CYCLIZATION REACTION WHICH CONVERTS
CC LYCOPENE TO BETA-CAROTENE AND NEUROSPORENE TO BETA-ZEACAROTENE.
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: BELONGS TO THE LYCOPENE CYCLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X81787; CAA57386.1; -.
DR PIR; S72506; S72506.
DR InterPro; IPR000205; NAD_binding.
KW Oxidoreductase; NAD; Carotenoid biosynthesis; Chloroplast;
KW Transit peptide.
FT TRANSIT 1 81 CHLOROPLAST (POTENTIAL).

FT CHAIN 82 500 LYCOPENE BETA CYCLASE.
 FT NP_BIND 86 114 NAD (POTENTIAL).
 SQ SEQUENCE 500 AA; 56067 MW; 2E3721B87EE6CBC8 CRC64;

Query Match 58.2%; Score 39; DB 1; Length 500;
 Best Local Similarity 66.7%; Pred. No. 50;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CENWWGDVC 9
 |||| ||
 Db 41 CENWGKGVVC 49

RESULT 10

ENV_HV1EL
 ID ENV_HV1EL STANDARD; PRT; 853 AA.
 AC P04581;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Envelope polyprotein GP160 precursor [Contains: Exterior membrane
 DE glycoprotein (GP120); Transmembrane glycoprotein (GP41)].
 GN ENV.
 OS Human immunodeficiency virus type 1 (ELI isolate) (HIV-1).
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86245056; PubMed=2424612;
 RA Alizon M., Wain-Hobson S., Montagnier L., Sonigo P.;
 RT "Genetic variability of the AIDS virus: nucleotide sequence analysis
 RT of two isolates from African patients."
 RL Cell 46:63-74(1986).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; K03454; AAA44329.1; -.
 DR EMBL; A07108; CAA00616.1; -.
 DR HIV; K03454; ENV\$ELI.
 DR InterPro; IPR000328; Env_GP41.
 DR InterPro; IPR000777; GP120.
 DR Pfam; PF00516; GP120; 1.
 DR Pfam; PF00517; GP41; 1.
 KW AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;
 KW Signal.
 FT SIGNAL 1 31 BY SIMILARITY.
 FT CHAIN 32 508 EXTERIOR MEMBRANE GLYCOPROTEIN.
 FT CHAIN 509 853 TRANSMEMBRANE GLYCOPROTEIN.
 FT DISULFID 53 73 BY SIMILARITY.
 FT DISULFID 118 206 BY SIMILARITY.
 FT DISULFID 125 197 BY SIMILARITY.

FT	DISULFID	130	154	BY SIMILARITY.
FT	DISULFID	219	248	BY SIMILARITY.
FT	DISULFID	229	240	BY SIMILARITY.
FT	DISULFID	297	330	BY SIMILARITY.
FT	DISULFID	376	442	BY SIMILARITY.
FT	DISULFID	383	416	BY SIMILARITY.
FT	CARBOHYD	87	87	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	129	129	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	137	137	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	143	143	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	153	153	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	157	157	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	183	183	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	188	188	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	198	198	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	235	235	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	242	242	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	263	263	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	277	277	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	290	290	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	331	331	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	353	353	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	384	384	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	390	390	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	394	394	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	400	400	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	405	405	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	406	406	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	411	411	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	445	445	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	458	458	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	459	459	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	462	462	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	608	608	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	613	613	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	622	622	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	634	634	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	853 AA; 96721 MW; F9CD864DAA0D07A5 CRC64;		

Query Match 58.2%; Score 39; DB 1; Length 853;
 Best Local Similarity 80.0%; Pred. No. 83;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CENWW 5
 |:|||
 Db 10 CQNWW 14

RESULT 11
 ADP1_YEAST
 ID ADP1_YEAST STANDARD; PRT; 1049 AA.
 AC P25371;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable ATP-dependent permease precursor.
 GN ADP1 OR YCR011C OR YCR11C OR YCR105.

OS *Saccharomyces cerevisiae* (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; *Saccharomyces*.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92160395; PubMed=1789009;
 RA Purnelle B., Skala J., Goffeau A.;
 RT "The product of the YCR105 gene located on the chromosome III from
 RT *Saccharomyces cerevisiae* presents homologies to ATP-dependent
 RT permeases.";
 RL Yeast 7:867-872(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92327849; PubMed=1626432;
 RA Skala J., Purnelle B., Goffeau A.;
 RT "The complete sequence of a 10.8 kb segment distal of *SUF2* on the
 RT right arm of chromosome III from *Saccharomyces cerevisiae* reveals
 RT seven open reading frames including the *RVS161*, *ADP1* and *PGK* genes.";
 RL Yeast 8:409-417(1992).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X59720; CAA42328.1; -.
 DR PIR; S19421; S19421.
 DR SGD; S0000604; ADP1.
 DR GO; GO:0005783; C:endoplasmic reticulum; IDA.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 KW ATP-binding; Transmembrane; Glycoprotein; Transport; Signal.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 1049 PROBABLE ATP-DEPENDENT PERMEASE.
 FT NP_BIND 423 430 ATP (BY SIMILARITY).
 FT TRANSMEM 325 345 POTENTIAL.
 FT TRANSMEM 464 481 POTENTIAL.
 FT TRANSMEM 794 814 POTENTIAL.
 FT TRANSMEM 829 849 POTENTIAL.
 FT TRANSMEM 878 898 POTENTIAL.
 FT TRANSMEM 910 930 POTENTIAL.
 FT TRANSMEM 938 958 POTENTIAL.
 FT TRANSMEM 1001 1021 POTENTIAL.
 FT TRANSMEM 1025 1045 POTENTIAL.
 FT CARBOHYD 50 50 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 815 815 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 935 935 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 960 960 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 971 971 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1049 AA; 117231 MW; ABC9CE54BCFDF6A3 CRC64;

Query Match 58.2%; Score 39; DB 1; Length 1049;
 Best Local Similarity 53.8%; Pred. No. 1e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 4; Gaps 1;

Qy 1 CENWWG----DVC 9
 |:| || |||
 Db 119 CDNGWGGINCDVC 131

RESULT 12

EX5C_HAEIN

ID EX5C_HAEIN STANDARD; PRT; 1121 AA.
 AC P44945;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Exodeoxyribonuclease V gamma chain (EC 3.1.11.5).
 GN RECC OR HI0942.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT Rd.";
 RL Science 269:496-512(1995).
 CC -!- FUNCTION: EXHIBITS A WIDE VARIETY OF CATALYTIC ACTIVITIES
 CC INCLUDING ATP-DEPENDENT EXONUCLEASE, ATP-STIMULATED ENDONUCLEASE,
 CC ATP-DEPENDENT HELICASE AND DNA-DEPENDENT ATPASE ACTIVITIES
 CC (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage (in the presence of
 CC ATP) in either 5'- to 3'-or 3'- to 5'-direction to yield 5'-
 CC phosphooligonucleotides.
 CC -!- SUBUNIT: CONSIST OF THREE SUBUNITS; RECB, RECC AND RECD
 CC (BY SIMILARITY).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U32775; AAC22596.1; -.
DR PIR; G64103; G64103.
DR TIGR; HI0942; -.
DR InterPro; IPR006347; ExoDNase_Vg.
DR Pfam; PF04257; Exonuc_V_gamma; 1.
DR TIGRFAMs; TIGR01450; recC; 1.
KW Hydrolase; Nuclease; Exonuclease; Endonuclease; Helicase; DNA repair;
KW Complete proteome.
SQ SEQUENCE 1121 AA; 129668 MW; E5070957296AE0D3 CRC64;

Query Match 58.2%; Score 39; DB 1; Length 1121;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CENWWGDV 8
|: :|||:
Db 263 CQEYWGDI 270

RESULT 13

FOL1_BOVIN

ID FOL1_BOVIN STANDARD; PRT; 222 AA.
AC P02702;
DT 21-JUL-1986 (Rel. 01, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Milk folate-binding protein (FBP) (Folate receptor alpha).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE.
RC TISSUE=Milk;
RA Svendsen I., Hansen S.I., Holm J., Lyngbye J.;
RT "The complete amino acid sequence of the folate-binding protein from
RT cow's milk.";
RL Carlsberg Res. Commun. 49:123-131(1984).
RN [2]
RP SEQUENCE OF 1-62; 72-102 AND 192-222.
RC TISSUE=Milk;
RA Svendsen I., Martin B., Pedersen T.G., Hansen S.I., Holm J.,
RA Lyngbye J.;
RT "Isolation and characterization of the folate-binding protein from
RT cow's milk.";
RL Carlsberg Res. Commun. 44:89-99(1979).
CC -!- FUNCTION: BINDS TO FOLATE AND REDUCED FOLIC ACID DERIVATIVES AND
CC MEDIATES DELIVERY OF 5-METHYLTETRAHYDROFOLATE TO THE INTERIOR OF
CC CELLS.

CC -!- PTM: EIGHT DISULFIDE BONDS ARE PRESENT.
 CC -!- SIMILARITY: BELONGS TO THE FOLATE RECEPTOR FAMILY.
 DR PIR; A03161; BFBO.
 DR InterPro; IPR004269; Folate_rec.
 DR Pfam; PF03024; Folate_rec; 1.
 KW Receptor; Glycoprotein; Milk; Folate-binding.
 FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .).
 SQ SEQUENCE 222 AA; 25825 MW; 528C388E9A9C0484 CRC64;

Query Match 56.7%; Score 38; DB 1; Length 222;
 Best Local Similarity 57.1%; Pred. No. 32;
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CENWWGD 7
 |::|||
 Db 119 CQSWWED 125

RESULT 14

CYA8_RAT

ID CYA8_RAT STANDARD; PRT; 1248 AA.
 AC P40146;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Adenylate cyclase, type VIII (EC 4.6.1.1) (ATP pyrophosphate-lyase)
 DE (Ca(2+)/calmodulin activated adenylyl cyclase).
 GN ADCY8.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;
 RX MEDLINE=94216337; PubMed=8163524;
 RA Cali J.J., Zwaagstra J.C., Mons N., Cooper D.M., Krupinski J.;
 RT "Type VIII adenylyl cyclase. A Ca²⁺/calmodulin-stimulated enzyme
 RT expressed in discrete regions of rat brain."
 RL J. Biol. Chem. 269:12190-12195(1994).
 CC -!- FUNCTION: This is a membrane-bound, calcium-inhibitable adenylyl
 CC cyclase. May be involved in learning, in memory and in drug
 CC dependence.
 CC -!- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
 CC -!- COFACTOR: Binds 2 magnesium ions per subunit (By similarity).
 CC -!- ENZYME REGULATION: Activated by calcium/calmodulin.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: BRAIN.
 CC -!- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
 CC -!- SIMILARITY: Belongs to the adenylyl cyclase class-4/guanylyl
 CC cyclase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; L26986; AAA20504.1; -.
 DR PIR; A53588; A53588.
 DR HSSP; P19754; lAWK.
 DR InterPro; IPR001054; G_cyclase.
 DR Pfam; PF00211; guanylate_cyc; 2.
 DR SMART; SM00044; CYCc; 2.
 DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 2.
 DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 2.
 KW Lyase; cAMP biosynthesis; Transmembrane; Glycoprotein; Repeat;
 KW Metal-binding; Magnesium.
 FT DOMAIN 1 179 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 180 200 POTENTIAL.
 FT TRANSMEM 209 229 POTENTIAL.
 FT TRANSMEM 244 264 POTENTIAL.
 FT TRANSMEM 271 291 POTENTIAL.
 FT TRANSMEM 293 313 POTENTIAL.
 FT TRANSMEM 318 338 POTENTIAL.
 FT DOMAIN 339 712 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 713 733 POTENTIAL.
 FT TRANSMEM 735 755 POTENTIAL.
 FT TRANSMEM 784 804 POTENTIAL.
 FT TRANSMEM 828 848 POTENTIAL.
 FT TRANSMEM 858 878 POTENTIAL.
 FT TRANSMEM 891 911 POTENTIAL.
 FT DOMAIN 912 1248 CYTOPLASMIC (POTENTIAL).
 FT METAL 416 416 MAGNESIUM 1 AND 2 (BY SIMILARITY).
 FT METAL 417 417 MAGNESIUM 2 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 460 460 MAGNESIUM 1 AND 2 (BY SIMILARITY).
 FT CARBOHYD 814 814 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 818 818 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 885 885 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1248 AA; 139822 MW; 0171A3CEED034961 CRC64;

Query Match 56.7%; Score 38; DB 1; Length 1248;
 Best Local Similarity 55.6%; Pred. No. 1.7e+02;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CENWWGDVC 9
 ||: || :|
 Db 1050 CEDKWGHLC 1058

RESULT 15
 CYA8_MOUSE
 ID CYA8_MOUSE STANDARD; PRT; 1249 AA.
 AC P97490;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Adenylate cyclase, type VIII (EC 4.6.1.1) (ATP pyrophosphate-lyase)
 DE (Ca(2+)/calmodulin activated adenyl cyclase).
 GN ADCY8.

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Brain;
RA Premont R.T.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: This is a membrane-bound, calcium-inhibitable adenylyl
CC cyclase. May be involved in learning, in memory and in drug
CC dependence.
CC -!- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
CC -!- COFACTOR: Binds 2 magnesium ions per subunit (By similarity).
CC -!- ENZYME REGULATION: Activated by calcium/calmodulin.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC -!- SIMILARITY: Belongs to the adenylyl cyclase class-4/guanylyl
CC cyclase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U85021; AAB41885.1; -.
DR HSSP; P19754; 1AWK.
DR MGD; MGI:1341110; Adcy8.
DR InterPro; IPR001054; G_cyclase.
DR Pfam; PF00211; guanylate_cyc; 2.
DR SMART; SM00044; CYCc; 2.
DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 2.
DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 2.
KW Lyase; cAMP biosynthesis; Transmembrane; Glycoprotein; Repeat;
KW Metal-binding; Magnesium.
FT DOMAIN 1 180 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 181 201 POTENTIAL.
FT TRANSMEM 210 230 POTENTIAL.
FT TRANSMEM 245 265 POTENTIAL.
FT TRANSMEM 272 292 POTENTIAL.
FT TRANSMEM 294 314 POTENTIAL.
FT TRANSMEM 319 339 POTENTIAL.
FT DOMAIN 340 713 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 714 734 POTENTIAL.
FT TRANSMEM 736 756 POTENTIAL.
FT TRANSMEM 785 805 POTENTIAL.
FT TRANSMEM 829 849 POTENTIAL.
FT TRANSMEM 859 879 POTENTIAL.
FT TRANSMEM 892 912 POTENTIAL.
FT DOMAIN 913 1249 CYTOPLASMIC (POTENTIAL).
FT METAL 417 417 MAGNESIUM 1 AND 2 (BY SIMILARITY).
FT METAL 418 418 MAGNESIUM 2 (VIA CARBONYL OXYGEN) (BY
FT SIMILARITY).
FT METAL 461 461 MAGNESIUM 1 AND 2 (BY SIMILARITY).

```

FT CARBOHYD 815 815 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 819 819 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 886 886 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1249 AA; 140154 MW; B2FE5670E9A74DAF CRC64;

Query Match 56.7%; Score 38; DB 1; Length 1249;
Best Local Similarity 55.6%; Pred. No. 1.7e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CENWWGDVC 9
||: ||: |
Db 1051 CEDKWHLC 1059

Search completed: November 13, 2003, 09:46:31
Job time : 7.15625 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 23.7188 Seconds
(without alignments)
97.917 Million cell updates/sec

Title: US-09-228-866-2
Perfect score: 67
Sequence: 1 CENWWGDVC 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*

```

12:  sp_virus:*
13:  sp_vertebrate:*
14:  sp_unclassified:*
15:  sp_rvirus:*
16:  sp_bacteriap:*
17:  sp_archeap:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query	Match Length			
1	47	70.1	252	5	Q9VT15	Q9vt15 drosophila
2	44	65.7	267	5	Q21692	Q21692 caenorhabdi
3	44	65.7	1132	16	Q8YLA5	Q8yla5 anabaena sp
4	43	64.2	249	6	Q9XSH1	Q9xsh1 sus scrofa
5	42	62.7	1008	4	Q8IVQ8	Q8ivq8 homo sapien
6	41	61.2	240	3	O93877	O93877 fusarium ox
7	41	61.2	242	3	Q04701	Q04701 fusarium so
8	41	61.2	244	3	Q00851	Q00851 nectria hae
9	41	61.2	333	6	Q9GKT2	Q9gkt2 macaca fasc
10	41	61.2	333	6	Q9BGQ4	Q9bgq4 macaca fasc
11	41	61.2	401	5	Q21938	Q21938 caenorhabdi
12	41	61.2	425	17	Q8PVY9	Q8pvy9 methanosarc
13	41	61.2	735	16	Q8PPY7	Q8ppy7 xanthomonas
14	41	61.2	744	17	Q8TJY7	Q8tjy7 methanosarc
15	40	59.7	85	4	O14597	O14597 homo sapien
16	40	59.7	274	13	Q9PW81	Q9pw81 gallus gall
17	40	59.7	337	2	Q93LL2	Q93ll2 nostoc linc
18	40	59.7	425	16	O25142	O25142 helicobacte
19	40	59.7	430	5	Q8IMC1	Q8imc1 drosophila
20	40	59.7	476	16	O25366	O25366 helicobacte
21	40	59.7	478	2	O30511	O30511 helicobacte
22	40	59.7	501	5	Q9GNS9	Q9gns9 trypanosoma
23	40	59.7	572	5	Q9V4E7	Q9v4e7 drosophila
24	40	59.7	582	16	Q9Z892	Q9z892 chlamydia p
25	40	59.7	629	16	Q9JS50	Q9js50 chlamydia p
26	40	59.7	685	4	Q8IWK5	Q8iwk5 homo sapien
27	40	59.7	718	3	O94277	O94277 schizosacch
28	40	59.7	767	10	Q8GSZ3	Q8gsz3 oryza sativ
29	40	59.7	799	10	Q9ZTJ7	Q9ztj7 lycopersico
30	40	59.7	826	2	Q8KL08	Q8kl08 rhizobium e
31	40	59.7	861	10	Q9SLS3	Q9sls3 nicotiana t
32	40	59.7	884	16	Q8P5B4	Q8p5b4 xanthomonas
33	40	59.7	926	4	Q8TE49	Q8te49 homo sapien
34	40	59.7	926	11	Q8R554	Q8r554 mus musculu
35	40	59.7	944	10	Q9ZTJ9	Q9ztj9 lycopersico
36	40	59.7	968	10	Q9ZTK1	Q9ztk1 lycopersico
37	40	59.7	1016	10	Q9ZTJ6	Q9ztj6 lycopersico
38	40	59.7	1112	10	Q41397	Q41397 lycopersico
39	40	59.7	1112	10	Q41398	Q41398 lycopersico
40	40	59.7	1148	16	Q9KPP4	Q9kpp4 vibrio chol
41	40	59.7	1270	16	Q8EF44	Q8ef44 shewanella

42	39	58.2	74	17	O29430	O29430 archaeoglob
43	39	58.2	124	5	Q8MN86	Q8mn86 dictyosteli
44	39	58.2	207	10	Q9LE73	Q9le73 arabidopsis
45	39	58.2	208	8	Q9G9G1	Q9g9g1 lasaea sp.

ALIGNMENTS

RESULT 1

Q9VT15

ID Q9VT15 PRELIMINARY; PRT; 252 AA.
AC Q9VT15;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG3088 protein (GH14734p).
GN CG3088.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
 RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celniker S.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 DR EMBL; AE003551; AAF50241.1; -.
 DR EMBL; AY094713; AAM11066.1; -.
 DR HSSP; P00766; 1GCT.
 DR FlyBase; FBgn0036015; CG3088.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001284; Ribosomal_L34E.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SpC; 1.
 DR PROSITE; PS01145; RIBOSOMAL_L34E; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 KW Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 252 AA; 26882 MW; 72F94448CF41DCEB CRC64;

Query Match 70.1%; Score 47; DB 5; Length 252;
 Best Local Similarity 75.0%; Pred. No. 5.9;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ENWWGDVC 9
 |||| :||
 Db 138 ENWWANVC 145

RESULT 2

Q21692

ID Q21692 PRELIMINARY; PRT; 267 AA.
 AC Q21692;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical 31.4 kDa protein.
 GN R04A9.3.
 OS *Caenorhabditis elegans*.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; *Caenorhabditis*.
 OX NCBI_TaxID=6239;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Geisel C.;
 RT "The sequence of C. elegans cosmid R04A9.";
 RL Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RT "Direct Submission.";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; U41550; AAA83285.1; -.
 DR WormPep; R04A9.3; CE04790.
 KW Hypothetical protein.
 SQ SEQUENCE 267 AA; 31377 MW; B951DAF6D7EA3A0B CRC64;

Query Match 65.7%; Score 44; DB 5; Length 267;
 Best Local Similarity 62.5%; Pred. No. 19;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CENWWGDV 8
 |: ||||:
 Db 19 CQAWWGDL 26

RESULT 3

Q8YLA5

ID Q8YLA5 PRELIMINARY; PRT; 1132 AA.
 AC Q8YLA5;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Hypothetical protein All7030.
 GN ALL7030.
 OS Anabaena sp. (strain PCC 7120).
 OG Plasmid pCC7120alpha.
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 RT cyanobacterium Anabaena sp. strain PCC 7120.";
 RL DNA Res. 8:205-213(2001).

DR EMBL; AP003600; BAB78114.1; -.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 1132 AA; 128203 MW; 236D2A33694B36F1 CRC64;

Query Match 65.7%; Score 44; DB 16; Length 1132;
Best Local Similarity 62.5%; Pred. No. 80;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CENWWGDV 8
|::||| |
Db 1065 CDSWWGQV 1072

RESULT 4

Q9XSH1

ID Q9XSH1 PRELIMINARY; PRT; 249 AA.
AC Q9XSH1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Membrane-bound folate binding protein.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Vallet J.L., Smith T.P.L., Sontegard T., Pearson P.L.,
RA Christenson R.K., Klemcke H.G.;
RT "Isolation of cDNAs encoding putative secreted and membrane-bound
RT folate binding proteins from endometrium of swine.";
RL Biol. Reprod. 0:0-0(1999).
DR EMBL; AF137374; AAD33741.1; -.
DR InterPro; IPR004269; Folate_rec.
DR Pfam; PF03024; Folate_rec; 1.
SQ SEQUENCE 249 AA; 28755 MW; 17FAAF2001D6B420 CRC64;

Query Match 64.2%; Score 43; DB 6; Length 249;
Best Local Similarity 71.4%; Pred. No. 25;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CENWWGD 7
|:|||| |
Db 131 CQNWWED 137

RESULT 5

Q8IVQ8

ID Q8IVQ8 PRELIMINARY; PRT; 1008 AA.
AC Q8IVQ8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Likely ortholog of mouse exportin 4 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC042504; AAH42504.1; -.
FT NON_TER 1 1
SQ SEQUENCE 1008 AA; 114035 MW; 9C94B2251B974494 CRC64;

Query Match 62.7%; Score 42; DB 4; Length 1008;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CENWW 5
|||
Db 580 CENWW 584

RESULT 6

093877

ID 093877 PRELIMINARY; PRT; 240 AA.
AC 093877;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Pectate lyase.
OS Fusarium oxysporum f. sp. lycopersici.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; mitosporic Hypocreales; Fusarium.
OX NCBI_TaxID=59765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=42-87;
RX MEDLINE=99146871; PubMed=10022947;
RA Huertas-Gonzalez M.D., Ruiz-Roldan M.C., Garcia Maceira F.I.,
RA Roncero M.I., Di Pietro A.;
RT "Cloning and characterization of pl1 encoding an in planta-secreted
RT pectate lyase of Fusarium oxysporum.";
RL Curr. Genet. 35:36-40(1999).
DR EMBL; AF080485; AAC64368.1; -.
DR InterPro; IPR004898; Pect_lyase.
DR Pfam; PF03211; Pectate_lyase; 1.
KW Lyase.
SQ SEQUENCE 240 AA; 24859 MW; 46D4B297305006B1 CRC64;

Query Match 61.2%; Score 41; DB 3; Length 240;
Best Local Similarity 83.3%; Pred. No. 51;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 WWGDVC 9
|||
Db 100 WWADVC 105

RESULT 7

Q04701

ID Q04701 PRELIMINARY; PRT; 242 AA.
AC Q04701;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Pectate lyase A precursor (EC 4.2.2.2).
GN PELA.
OS Fusarium solani (subsp. pisi).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Nectriaceae; Nectria.
OX NCBI_TaxID=70791;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 18-55 AND 69-88.
RX MEDLINE=93015682; PubMed=1400187;
RA Gonzalez-Candelas L., Kolattukudy P.E.;
RT "Isolation and analysis of a novel inducible pectate lyase gene from
RT the phytopathogenic fungus Fusarium solani f. sp. pisi (Nectria
RT haematococca, mating population VI).";
RL J. Bacteriol. 174:6343-6349(1992).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 21-25.
RX MEDLINE=88022783; PubMed=3310898;
RA Crawford M.S., Kolattukudy P.E.;
RT "Pectate lyase from Fusarium solani f. sp. pisi: purification,
RT characterization, in vitro translation of the mRNA, and involvement in
RT pathogenicity.";
RL Arch. Biochem. Biophys. 258:196-205(1987).
CC -!- FUNCTION: INVOLVED IN MACERATION AND SOFT-ROTTING OF PLANT TISSUE.
CC -!- CATALYTIC ACTIVITY: ELIMINATIVE CLEAVAGE OF PECTATE TO GIVE
CC OLIGOSACCHARIDES WITH 4-DEOXY-ALPHA-D-GLUC-4-ENURONOSYL GROUPS AT
CC THEIR NON-REDUCING ENDS.
CC -!- ENZYME REGULATION: SUBJECT TO SELF CATABOLITE REPRESSION.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- INDUCTION: BY PECTIN.
CC -!- DISEASE: PECTATE LYASES HAVE BEEN IMPLICATED AS PATHOGENICITY
CC FACTORS WHICH INDUCE MACERATION OR ROTTING OF PLANT TISSUE.
CC -!- SIMILARITY: BELONGS TO THE PLADES FAMILY OF EXTRACELLULAR PELS.
CC SIMILAR TO THE PLBC PROTEINS.
DR EMBL; M94691; AAA33338.1; -.
DR EMBL; M94692; AAA33339.1; -.
DR InterPro; IPR004898; Pect_lyase.
DR Pfam; PF03211; Pectate_lyase; 1.
KW Lyase; Multigene family; Signal; Glycoprotein.
FT SIGNAL 1 17
FT CHAIN 18 242 PECTATE LYASE A.
FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 242 AA; 25339 MW; 3F338FBE895AB286 CRC64;

Query Match 61.2%; Score 41; DB 3; Length 242;
Best Local Similarity 83.3%; Pred. No. 52;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 WWGDVC 9
|| |||
Db 102 WWADVC 107

RESULT 8

Q00851

ID Q00851 PRELIMINARY; PRT; 244 AA.
AC Q00851;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Pectate lyase B.
GN PELB.
OS Nectria haematococca.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Nectriaceae; Nectria.
OX NCBI_TaxID=140110;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=T8;
RX MEDLINE=96099288; PubMed=8522511;
RA Guo W., Gonzalez-Candelas L., Kolattukudy P.E.;
RT "Cloning of a novel constitutively expressed pectate lyase gene pelB
RT from Fusarium solani f. sp. pisi (Nectria haematococca, mating type
RT VI) and characterization of the gene product expressed in Pichia
RT pastoris.";
RL J. Bacteriol. 177:7070-7077(1995).
DR EMBL; U13051; AAA87383.1; -.
DR InterPro; IPR004898; Pect_lyase.
DR Pfam; PF03211; Pectate_lyase; 1.
KW Lyase.
SQ SEQUENCE 244 AA; 25663 MW; BF804413A6546469 CRC64;

Query Match 61.2%; Score 41; DB 3; Length 244;
Best Local Similarity 83.3%; Pred. No. 52;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 WWGDVC 9
|| |||
Db 103 WWADVC 108

RESULT 9

Q9GKT2

ID Q9GKT2 PRELIMINARY; PRT; 333 AA.
AC Q9GKT2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical 38.9 kDa protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;

RT "Isolation of full-length cDNA clones from macaque brain cDNA
 RT libraries.";
 RL Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AB052147; BAB19002.1; -.
 DR InterPro; IPR006087; Sterol_desat.
 DR InterPro; IPR006088; Sterol_desatur.
 DR Pfam; PF01598; Sterol_desat; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 333 AA; 38885 MW; B99755D843312A06 CRC64;

Query Match 61.2%; Score 41; DB 6; Length 333;
 Best Local Similarity 83.3%; Pred. No. 71;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 WWGDVC 9
 |||||
 Db 156 WWGDPC 161

RESULT 10

Q9BGQ4

ID Q9BGQ4 PRELIMINARY; PRT; 333 AA.
 AC Q9BGQ4;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Hypothetical 38.9 kDa protein.
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
 OC Cercopithecinae; Macaca.
 OX NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Frontal cortex;
 RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
 RA Suzuki Y., Sugano S., Hashimoto K.;
 RT "Isolation of full-length cDNA clones from macaque brain cDNA
 RT libraries.";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AB056418; BAB33076.1; -.
 DR InterPro; IPR006087; Sterol_desat.
 DR InterPro; IPR006088; Sterol_desatur.
 DR Pfam; PF01598; Sterol_desat; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 333 AA; 38925 MW; 30795B28433138B1 CRC64;

Query Match 61.2%; Score 41; DB 6; Length 333;
 Best Local Similarity 83.3%; Pred. No. 71;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 WWGDVC 9
 |||||
 Db 156 WWGDPC 161

RESULT 11

Q21938

ID Q21938 PRELIMINARY; PRT; 401 AA.
AC Q21938;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE R11D1.10 protein.
GN R11D1.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Steward C.A.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z75547; CAA99905.3; -.
DR WormPep; R11D1.10; CE28948.
DR InterPro; IPR000306; Znf_FYVE.
DR SMART; SM00064; FYVE; 1.
DR PROSITE; PS50178; ZF_FYVE; 1.
SQ SEQUENCE 401 AA; 46304 MW; FE45918C5DB1E755 CRC64;

Query Match 61.2%; Score 41; DB 5; Length 401;
Best Local Similarity 83.3%; Pred. No. 86;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ENWWGD 7
:|||||
Db 91 KNWWGD 96

RESULT 12

Q8PVY9

ID Q8PVY9 PRELIMINARY; PRT; 425 AA.
AC Q8PVY9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Conserved protein.
GN MM1816.
OS Methanosarcina mazei (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2209;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Goe1 / Go1 / ATCC BAA-199 / DSM 3647 / OCM 88;
RX MEDLINE=22120827; PubMed=12125824;
RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,

RA Martinez-Arias R., Henne A., Wiezer A., Baeumer S., Jacobi C.,
 RA Brueggemann H., Lienard T., Christmann A., Boemecke M., Steckel S.,
 RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
 RA Fritz H.-J., Gottschalk G.;
 RT "The genome of *Methanosarcina mazei*: evidence for lateral gene
 RT transfer between Bacteria and Archaea.";
 RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
 DR EMBL; AE013418; AAM31512.1; -.
 DR InterPro; IPR006457; S_layer_rel_Mac.
 DR TIGRFAMS; TIGR01567; S_layer_rel_Mac; 1.
 KW Complete proteome.
 SQ SEQUENCE 425 AA; 46508 MW; 4826E1542DF69D97 CRC64;

Query Match 61.2%; Score 41; DB 17; Length 425;
 Best Local Similarity 83.3%; Pred. No. 91;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ENWWGD 7
 |||||:
 Db 71 ENWWGE 76

RESULT 13

Q8PPY7

ID Q8PPY7 PRELIMINARY; PRT; 735 AA.
 AC Q8PPY7;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Hypothetical protein XAC0546.
 GN XAC0546.
 OS *Xanthomonas axonopodis* (pv. citri).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=92829;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=306 / ATCC 13902 / XV 101;
 RX MEDLINE=22022145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
 RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
 RA Setubal J.C., Kitajima J.P.;
 RT "Comparison of the genomes of two *Xanthomonas* pathogens with differing
 RT host specificities.";
 RL Nature 417:459-463(2002).

DR EMBL; AE011681; AAM35435.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 735 AA; 80474 MW; 8D5E23E1DF991D69 CRC64;

Query Match 61.2%; Score 41; DB 16; Length 735;
Best Local Similarity 83.3%; Pred. No. 1.6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 NWWGDV 8
|||||:
Db 593 NWWGDL 598

RESULT 14

Q8TJY7

ID Q8TJY7 PRELIMINARY; PRT; 744 AA.
AC Q8TJY7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Predicted protein.
GN MA3639.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL; AE011072; AAM06994.1; -.
DR InterPro; IPR006457; S_layer_rel_Mac.
DR TIGRFAMS; TIGR01567; S_layer_rel_Mac; 2.
KW Complete proteome.
SQ SEQUENCE 744 AA; 83248 MW; 3E822964E71D5C04 CRC64;

Query Match 61.2%; Score 41; DB 17; Length 744;
Best Local Similarity 83.3%; Pred. No. 1.6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ENWWGD 7
|||||:
Db 389 ENWWGE 394

RESULT 15

014597

ID 014597 PRELIMINARY; PRT; 85 AA.
 AC 014597;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Folate binding protein (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Salivary gland;
 RA Verma R.S., Elwood P.C.;
 RT "Identification and characterization of homologous cDNA to KB folate
 RT receptor from human salivary gland."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AF000381; AAB81938.2; -.
 DR InterPro; IPR004269; Folate_rec.
 DR Pfam; PF03024; Folate_rec; 1.
 FT NON_TER 1 1
 FT NON_TER 85 85
 SQ SEQUENCE 85 AA; 9995 MW; E39A8454EE396A76 CRC64;

Query Match 59.7%; Score 40; DB 4; Length 85;
 Best Local Similarity 71.4%; Pred. No. 26;
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CENWWGD 7
 || || |
 Db 36 CEQWWED 42

Search completed: November 13, 2003, 09:50:57
 Job time : 25.7188 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 30.2812 Seconds
(without alignments)
47.176 Million cell updates/sec

Title: US-09-228-866-3
Perfect score: 49
Sequence: 1 CLSSRLDAC 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result		%					
No.	Score	Query	Match	Length	DB	ID	Description
1	49	100.0		9	18	AAW13412	Brain homing pepti
2	49	100.0		9	21	AAB07389	Brain homing pepti
3	49	100.0		9	22	AAE11795	Phage peptide #3 t
4	49	100.0		9	23	AAU10706	Brain homing pepti
5	49	100.0		9	24	ABU59532	Brain receptor tar
6	40	81.6		20	21	AAB26819	Peptidic membrane
7	38	77.6		130	19	AAW77731	Exou protein. Sta
8	37	75.5		9	23	ABG60523	Selective targetin
9	37	75.5		9	23	ABG35135	Pancreatic islet t
10	36	73.5		62	22	AAM83339	Human immune/haema
11	36	73.5		209	22	AAU29476	Human G protein-co
12	36	73.5		209	23	ABG60764	Novel G protein co
13	35	71.4		53	20	AAU01191	Polypeptide fragme
14	35	71.4		80	22	AAU61079	Propionibacterium
15	35	71.4		87	22	AAM06519	Human foetal prote
16	35	71.4		255	21	AAB42698	Human ORFX ORF2462
17	35	71.4		257	22	ABG04734	Novel human diagno
18	34	69.4		54	21	AAB32402	Human secreted pro
19	34	69.4		96	21	AAB32483	Human secreted pro
20	34	69.4		100	23	AAM50063	RanBP 11-binding p
21	34	69.4		114	21	AAB08908	Human secreted pro
22	34	69.4		132	21	AAB08942	Human secreted pro
23	34	69.4		203	22	AAB74336	Human papillomavir
24	34	69.4		458	24	ABP57706	Saccharopolyspora
25	34	69.4		781	17	AAW01871	Recombinase enzyme
26	34	69.4		1147	22	AAU01207	Human caspase recr
27	34	69.4		1147	23	AAG79555	Human CARD-11. Ho
28	34	69.4		1247	24	ABG76061	Human caspase recr
29	33	67.3		65	23	AAU76371	Human airway tryps
30	33	67.3		118	23	ABG70785	E. coli RscC recei
31	33	67.3		167	22	ABG13149	Novel human diagno
32	33	67.3		175	22	AAU15924	Human novel secret
33	33	67.3		175	24	ABU54993	Human novel polype
34	33	67.3		309	13	AAR29642	pCTD ORF 7. Chlam
35	33	67.3		382	22	ABG03207	Novel human diagno
36	33	67.3		478	24	ABP97392	Mouse embryonic st
37	33	67.3		568	22	ABG01439	Novel human diagno
38	33	67.3		617	21	AAU76048	Murine skin cell p
39	33	67.3		617	22	AAB55987	Skin cell protein,
40	33	67.3		617	23	ABB72187	Murine protein iso
41	32	65.3		27	20	AAU25519	Human collectin pr
42	32	65.3		27	21	AAU77986	Modified consensus
43	32	65.3		27	22	AAG63349	Consensus sequence
44	32	65.3		45	22	AAO05612	Human polypeptide
45	32	65.3		111	22	AAU41226	Propionibacterium

ALIGNMENTS

RESULT 1
 AAW13412
 ID AAW13412 standard; Peptide; 9 AA.
 XX
 AC AAW13412;
 XX
 DT 15-JAN-1998 (first entry)
 XX
 DE Brain homing peptide.
 XX
 KW Brain homing peptide; in vivo panning; screening; phage display;
 KW drug delivery.
 XX
 OS Synthetic.
 XX
 PN WO9710507-A1.
 XX
 PD 20-MAR-1997.
 XX
 PF 10-SEP-1996; 96WO-US14600.
 XX
 PR 11-SEP-1995; 95US-0526710.
 PR 11-SEP-1995; 95US-0526708.
 XX
 PA (LJOL-) LA JOLLA CANCER RES FOUND.
 XX
 PI Pasqualini R, Ruoslahti E;
 XX
 DR WPI; 1997-202359/18.
 XX
 PT Obtaining compound that homes to selected organ or tissue - by in
 PT vivo panning method, specifically to identify brain, kidney,
 PT angiogenic vasculature or tumour tissue homing peptide(s)
 XX
 PS Claim 10; Page 67; 75pp; English.
 XX
 CC This synthetic peptide is a claimed example of a brain-homing
 CC peptide that was identified using a novel method for obtaining
 CC molecules that home to a selected organ or tissue. This in vivo
 CC panning method typically involves administering a phage display
 CC library to a subject, and identifying expressed peptides which
 CC home to the desired organ or tissue, e.g. brain, kidney, angiogenic
 CC vascular tissue or tumour tissue. The isolated peptides (see
 CC AAW13412-52, AAW11181-86) can be used to target e.g. drugs, toxins or
 CC labels to the selected organ/tissue (claimed) or to identify and/or
 CC isolate target molecules (claimed). The peptides can be directly
 CC identified in vivo, as compared to prior art in vitro screening
 CC methods, which require further examination to see if they maintain
 CC specificity in vivo.
 XX
 SQ Sequence 9 AA;

 Query Match 100.0%; Score 49; DB 18; Length 9;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 Qy 1 CLSSRLDAC 9

Db

|||||||
1 CLSSRLDAC 9

RESULT 2

AAB07389

ID AAB07389 standard; peptide; 9 AA.

XX

AC AAB07389;

XX

DT 17-OCT-2000 (first entry)

XX

DE Brain homing peptide # 3.

XX

KW Brain; homing peptide; organ targeting; tissue targeting; mouse; cyclic.

XX

OS Mus sp.

XX

FH Key Location/Qualifiers

FT Disulfide-bond 1..9

FT /note= "Can optionally form a cyclic peptide"

XX

PN US6068829-A.

XX

PD 30-MAY-2000.

XX

PF 23-JUN-1997; 97US-0862855.

XX

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Pasqualini R, Ruoslahti E;

XX

DR WPI; 2000-410850/35.

XX

PT Identifying and recovering organ homing molecules or peptides by in

PT vivo panning comprises administering a library of diverse peptides

PT linked to a tag which facilitates recovery of these peptides -

XX

PS Example 2; Column 17; 20pp; English.

XX

CC The present sequence is a mouse brain homing peptide. This sequence was

CC identified by using in vivo panning to screen a library of potential

CC organ homing molecules. The present sequence can be used to direct a

CC moiety to a the brain tissue, by linking the moiety to the present

CC sequence. Examples of potential moieties are drugs, toxins or a

CC detectable label. The present sequence contains a SRL amino acid motif.

XX

SQ Sequence 9 AA;

Query Match 100.0%; Score 49; DB 21; Length 9;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 CLSSRLDAC 9

Db

|||||||
1 CLSSRLDAC 9

RESULT 3

AAE11795

ID AAE11795 standard; peptide; 9 AA.

XX

AC AAE11795;

XX

DT 18-DEC-2001 (first entry)

XX

DE Phage peptide #3 targetted to brain.

XX

KW Enriched library fraction; brain; kidney; tumour; panning; diagnostic;

KW molecular medicine; drug delivery; peptidomimetic; pharmaceutical.

XX

OS Bacteriophage.

XX

FH Key Location/Qualifiers

FT Domain 4..6

FT /label= SRL_motif

XX

PN US6296832-B1.

XX

PD 02-OCT-2001.

XX

PF 08-JAN-1999; 99US-0226985.

XX

PR 23-JUN-1997; 97US-0862855.

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ruoslahti E, Pasqualini R;

XX

DR WPI; 2001-610691/70.

XX

PT Enriched library fraction comprising molecules recovered by in vivo
PT panning that selectively home to a selected organ or tissue useful for
PT treating disease or in diagnostic methods -

XX

PS Example 2; Column 17; 21pp; English.

XX

CC The invention relates to an enriched library fraction containing
CC molecules that selectively home to a selected organ or tissue such as
CC brain, kidney or tumour recovered by in vivo panning. The invention
CC generally relates to the field of molecular medicine, drug delivery and
CC to a method of invivo panning for identifying a molecule that homes to a
CC specific organ. The molecules, e.g., peptides, peptidomimetics, proteins
CC and fragments of proteins contained in an enriched library fraction may
CC be administered to a subject as part of a pharmaceutical composition to
CC treat disease or in diagnostic methods. The present sequence is a
CC peptide from bacteriophage targetted to brain.

XX

SQ Sequence 9 AA;

Query Match 100.0%; Score 49; DB 22; Length 9;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLSSRLDAC 9
|||
Db 1 CLSSRLDAC 9

RESULT 4

AAU10706

ID AAU10706 standard; peptide; 9 AA.

XX

AC AAU10706;

XX

DT 12-MAR-2002 (first entry)

XX

DE Brain homing peptide #3 useful for delivery of target molecules.

XX

KW Organ targeting; tissue targeting; cancer; tumour homing molecule;
KW delivery of target molecule; brain homing peptide.

XX

OS Synthetic.

XX

PN US6306365-B1.

XX

PD 23-OCT-2001.

XX

PF 08-JAN-1999; 99US-0227906.

XX

PR 23-JUN-1997; 97US-0862855.

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ruoslahti E, Pasqualini R;

XX

DR WPI; 2002-040196/05.

XX

PT Recovering molecules that home to an organ or tissue, useful for
PT identifying molecules that home to a specific organ or tissue, e.g.
PT identifying a tumour homing molecule to identify the presence of cancer,
PT by in vivo panning of a library -

XX

PS Example 2; Column 17; 21pp; English.

XX

CC The present invention relates to a method of recovering molecules that
CC home to a selected organ or tissue. The method comprises administering
CC to the subject the library of diverse molecules, collecting a sample of
CC the selected organ or tissue (e.g. brain or kidney), and recovering from
CC the sample several molecules that home to the selected organ or tissue.
CC The method is useful for identifying molecules, particularly useful for
CC screening large number of molecules (e.g. peptides), that home to a
CC specific organ. The identified molecule is useful for e.g. raising an
CC antibody specific for a target molecule, targeting a desired moiety

CC (e.g. drug, toxin or detectable label) to the selected organ.
CC Specifically, the method is useful for identifying the presence of cancer
CC in a subject by linking an appropriate moiety to a tumour homing
CC molecule. The present method provides a direct means for identifying
CC molecules that specifically home to a selected organ and, therefore
CC provides a significant advantage over previous methods, which require
CC that a molecule identified using an in vitro screening method
CC subsequently be examined to determine if it maintains its specificity in
CC vivo. AAU10704-AAU10723 represent brain homing peptides described in
CC the present invention.

XX

SQ Sequence 9 AA;

Query Match 100.0%; Score 49; DB 23; Length 9;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
|||
Db 1 CLSSRLDAC 9

RESULT 5

ABU59532

ID ABU59532 standard; Peptide; 9 AA.

XX

AC ABU59532;

XX

DT 22-APR-2003 (first entry)

XX

DE Brain receptor targeting peptide #4.

XX

KW Targeting ligand; bioactive agent; polymer matrix; cancer; cytostatic;
KW cathepsin-D substrate; peptides; neuroreceptor; adrenal receptor;
KW fibronectin; vitronectin; integrin; RGD motif; angiogenic endothelium;
KW tumour; cationic cancer-targeting peptide.

XX

OS Synthetic.

XX

PN US2002041898-A1.

XX

PD 11-APR-2002.

XX

PF 25-JUL-2001; 2001US-0912609.

XX

PR 05-JAN-2000; 2000US-0478124.

PR 31-OCT-2000; 2000US-0703474.

XX

PA (UNGE/) UNGER E C.

PA (MATS/) MATSUNAGA T O.

PA (RAMA/) RAMASWAMI V.

PA (ROMA/) ROMANOWSKI M J.

XX

PI Unger EC, Matsunaga TO, Ramaswami V, Romanowski MJ;

XX

DR WPI; 2003-208921/20.

XX

PT Targeted delivery system comprising a bioactive agent homogeneously
PT dispersed in a targeted matrix is especially useful in cancer therapy
PT -

XX

PS Claim 23; Page 37; 46pp; English.

XX

CC The invention relates to a composition comprising a bioactive agent
CC homogeneously dispersed in a targeted matrix (polymer and targeting
CC ligand). Also included are a targeted matrix for use as a delivery
CC vehicle comprising a polymer associated with a targeting ligand,
CC enhancing the bioavailability of an agent comprising administration
CC of the composition and treating cancer comprising administration of the
CC novel composition. The method is useful for targeted delivery of a drug,
CC especially in cancer therapy. The targeting ligand may be a peptide.
CC Examples of targeting peptides are disclosed including cathepsin-D
CC substrate peptides, peptides targeting receptors in the brain and
CC kidney, peptides recognising fibronectin- and vitronectin-binding
CC integrins, peptides targeting the RGD (Arg-Gly-Asp)-motif in, e.g.,
CC antibodies, peptides targeting the angiogenic endothelium of solid
CC tumours, tissue specific peptides (e.g. of lung, skin, pancreas,
CC intestine, uterus, adrenal gland and retina), and cationic cancer-
CC targeting peptides. The present sequence is a peptide targeting
CC ligand disclosed in the invention.

XX

SQ Sequence 9 AA;

Query Match 100.0%; Score 49; DB 24; Length 9;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
|||
Db 1 CLSSRLDAC 9

RESULT 6

AAB26819

ID AAB26819 standard; peptide; 20 AA.

XX

AC AAB26819;

XX

DT 23-JAN-2001 (first entry)

XX

DE Peptidic membrane binding element.

XX

KW Organ perfusion; transplantation; storage; antiinflammatory;
KW immunosuppressive; vasotropic; complement activation inhibitor;
KW allograft rejection; ischaemia reperfusion injury.

XX

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Modified-site 1

FT /note= "Optionally N-Myristoyl-Gly"

FT Modified-site 20

FT /note= "Optionally S-2-Thiopyridyl-Cys-NH2"

XX

PN WO200053007-A1.
 XX
 PD 14-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-GB00834.
 XX
 PR 10-MAR-1999; 99GB-0005503.
 XX
 PA (ADPR-) ADPROTECH LTD.
 XX
 PI Smith RAG, Pratt JR, Sacks SH;
 XX
 DR WPI; 2000-601920/57.
 XX
 PT Preparation for perfusing organ prior to transplantation or storage
 PT comprises soluble derivative of a soluble polypeptide which comprises
 PT two heterologous membrane binding elements with low membrane affinity
 PT -
 XX
 PS Example 2; Page 20; 47pp; English.
 XX
 CC The present invention relates to formulations and preparations for
 CC perfusing an organ prior to transplantation or storage. The preparation
 CC comprises a soluble derivative or a polypeptide, which has two or more
 CC heterologous membrane binding elements. The membrane binding elements are
 CC capable of interacting, independently and with thermodynamic additivity,
 CC with membrane components of the organ exposed to extracellular perfusion
 CC fluids, and a flush storage solution. The preparation exhibits
 CC antiinflammatory, immunosuppressive and vasotropic activity and works as
 CC a complement activation inhibitor and an inhibitor of cytotoxic T
 CC lymphocyte activity. The preparation is used for preparing an organ prior
 CC to transplantation or storage and for prevention, treatment or
 CC amelioration of a disease or disorder associated with inflammation,
 CC inappropriate complement activation or inappropriate activation of
 CC coagulant or thrombotic processes prior to, during or after
 CC transplantation or storage of an organ. The preparation is useful for
 CC treating hyperacute and acute allograft rejection of transplanted organs
 CC such as kidney, heart, liver or lungs, ischaemia-reperfusion injury in
 CC transplanted organs, xenograft rejection and corneal graft rejection. The
 CC present sequence represents a peptidic membrane binding element used in
 CC an example of the preparation of the invention.
 XX
 SQ Sequence 20 AA;

Query Match 81.6%; Score 40; DB 21; Length 20;
 Best Local Similarity 100.0%; Pred. No. 4.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LSSRLDAC 9
 |||||
 Db 13 LSSRLDAC 20

RESULT 7
 AAW77731
 ID AAW77731 standard; Protein; 130 AA.
 XX

AC AAW77731;
 XX
 DT 30-OCT-1998 (first entry)
 XX
 DE Exou protein.
 XX
 KW Staphylococcus aureus protein; immune response induction; eye infection;
 KW antibody production; T-cell immune response; gastrointestinal infection;
 KW respiratory infection; inhibitor; bacterial infection; cardiac infection;
 KW central nervous system; kidney infection; urinary tract infection;
 KW antimicrobial compound identification; broad spectrum antibiotic;
 KW therapy.
 XX
 OS Staphylococcus aureus.
 XX
 PN EP841394-A2.
 XX
 PD 13-MAY-1998.
 XX
 PF 24-SEP-1997; 97EP-0307485.
 XX
 PR 24-SEP-1996; 96US-0027032.
 XX
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 XX
 PI Black MT, Burnham MKR, Hodgson JE, Knowles DJC;
 PI Lonetto MA, Nicholas RO, Pratt JM, Reichard RW, Rosenberg M;
 PI Ward JM;
 XX
 DR WPI; 1998-252940/23.
 DR N-PSDB; AAV53520.
 XX
 PT New nucleic acid sequences from Staphylococcus aureus WCHU29 -
 PT useful in vaccines and for treatment of bacterial infections of e.g.
 PT respiratory tract and central nervous system
 XX
 PS Claim 11; Page 357-358; 390pp; English.
 XX
 CC This sequence represents a Staphylococcus aureus protein, that based on
 CC homology with a Rhizobium Meliloti protein, is a Exou protein,
 CC and is encoded by a DNA sequence of the invention.
 CC The DNA sequences were isolated from Staphylococcus aureus WCHU29
 CC (NCIMB 40771). Host cells containing the DNA sequences are used to
 CC produce polypeptides or fragments. The proteins are used in the treatment
 CC of disease, for inducing an immune response by administering them, to
 CC produce antibody and/or T-cell immune response. Antagonists of the
 CC proteins are used for the inhibition of bacterial polypeptides.
 CC Conditions which may be treated include bacterial infections, especially
 CC respiratory, cardiac, gastrointestinal, central nervous, eye, kidney,
 CC urinary tract, skin, bones and joints. The proteins can also be used to
 CC identify antimicrobial compounds which are broad spectrum antibiotics,
 CC especially useful in the treatment of H. pylori infection.
 XX
 SQ Sequence 130 AA;

Query Match

77.6%; Score 38; DB 19; Length 130;

Best Local Similarity 77.8%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
|| || ||
Db 112 CLLSRCDAC 120

RESULT 8

ABG60523

ID ABG60523 standard; Peptide; 9 AA.

XX

AC ABG60523;

XX

DT 30-JUL-2002 (first entry)

XX

DE Selective targeting peptide #198.

XX

KW Targeting peptide; cancer; arthritis; diabetes; inflammatory disease;
KW atherosclerosis; autoimmune disease; bacterial infection; apoptosis;
KW viral infection; cardiovascular disease; degenerative disease; ischaemia;
KW inflammation; macular degeneration; antiinflammatory; antidiabetic;
KW cardiovascular; immunomodulator; antibacterial; antiviral; cytostatic;
KW gene therapy.

XX

OS Synthetic.

XX

PN WO200220769-A1.

XX

PD 14-MAR-2002.

XX

PF 07-SEP-2001; 2001WO-US27692.

XX

PR 08-SEP-2000; 2000US-231266P.

PR 17-JAN-2001; 2001US-0765101.

XX

PA (TEXA) UNIV TEXAS SYSTEM.

XX

PI Arap W, Pasqualini R;

XX

DR WPI; 2002-415731/44.

XX

PT Targeting peptides identified by phage display, useful for targeting
PT delivery to an organ or tissue, particularly for treating a disease,
PT e.g. cancer, inflammatory or autoimmune diseases, infections or
PT cardiovascular disease -

XX

PS Claim 22; Page 121; 317pp; English.

XX

CC The invention relates to an isolated peptide of 100 amino acids or less
CC in size useful for targeting delivery to an organ or tissue, particularly
CC for treating a disease, e.g. cancer, arthritis, diabetes, inflammatory
CC disease, atherosclerosis, autoimmune disease, bacterial infection, viral
CC infection, cardiovascular disease or degenerative disease. The peptide is
CC also useful for inducing apoptosis, particularly to a subject with
CC ischaemia, cancer, arthritis, diabetes, cardiovascular disease,
CC inflammation or macular degeneration. Furthermore, the peptide is useful

CC for diagnosing the diseases cited above. Targeting peptides of the
CC invention can also be used to deliver an agent to a foetus, by attaching
CC a peptide to the agent and administering the peptide to a pregnant
CC subject. Sequences ABG60326-ABG60574 represent selective targeting
CC peptides of the invention.

XX

SQ Sequence 9 AA;

Query Match 75.5%; Score 37; DB 23; Length 9;
Best Local Similarity 66.7%; Pred. No. 9.3e+05;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9

||:| :|||

Db 1 CLASGMDAC 9

RESULT 9

ABG35135

ID ABG35135 standard; Peptide; 9 AA.

XX

AC ABG35135;

XX

DT 15-JUL-2002 (first entry)

XX

DE Pancreatic islet targeting peptide #11.

XX

KW Targeting peptide; cancer; Hodgkin's disease; cytostatic;

KW immunosuppressive; anti-inflammatory; antiarthritic; antiviral;

KW antiatherosclerotic; antidiabetic; antibacterial; diabetes mellitus;

KW inflammatory disease; arthritis; atherosclerosis; cancer;

KW autoimmune disease; bacterial infection; viral infection.

XX

OS Unidentified.

XX

PN WO200220722-A2.

XX

PD 14-MAR-2002.

XX

PF 07-SEP-2001; 2001WO-US27702.

XX

PR 08-SEP-2000; 2000US-231266P.

PR 17-JAN-2001; 2001US-0765101.

XX

PA (TEXA) UNIV TEXAS SYSTEM.

XX

PI Arap W, Pasqualini R;

XX

DR WPI; 2002-383050/41.

XX

PT Identifying targeting peptides useful for treating e.g. diabetes

PT mellitus, inflammatory diseases, cancer, or autoimmune diseases,

PT comprises exposing a sample to a phage display library and recovering

PT phage bound to the sample -

XX

PS Claim 56; Page 288; 298pp; English.

XX

CC This invention relates to a novel method for identifying disease
CC targeting peptides. The method comprises exposing a sample from an
CC organ, tissue or cell type of interest, to a phage display library and
CC recovering phage bound to the sample (the phage expresses targeting
CC peptides). The peptides identified by the method of the invention may
CC have cytostatic, immunosuppressive, anti-inflammatory, antiarthritic,
CC antiatherosclerotic, antidiabetic, antibacterial and antiviral
CC activities. The methods and composition are useful for identifying
CC targeting peptides and one or more receptors for a targeting peptide.
CC The targeting peptides are used for selective delivery of therapeutic
CC agents, including gene therapy vectors and fusion proteins, to specific
CC organs, tissues, or cell types in subject. The targeting peptide may
CC also be used for treating diseases such as diabetes mellitus,
CC inflammatory diseases, arthritis, atherosclerosis, cancer, autoimmune
CC diseases, bacterial and viral infections and Hodgkin's disease. The
CC present sequence represents a targeting peptide of the invention.

XX

SQ Sequence 9 AA;

Query Match 75.5%; Score 37; DB 23; Length 9;
Best Local Similarity 66.7%; Pred. No. 9.3e+05;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
||:| :|||
Db 1 CLASGMDAC 9

RESULT 10

AAM83339

ID AAM83339 standard; Protein; 62 AA.

XX

AC AAM83339;

XX

DT 07-NOV-2001 (first entry)

XX

DE Human immune/haematopoietic antigen SEQ ID NO:10932.

XX

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis.

XX

OS Homo sapiens.

XX

PN WO200157182-A2.

XX

PD 09-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US01354.

XX

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.

PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Barash SC, Ruben SM;

XX

DR WPI; 2001-483426/52.

DR N-PSDB; AAK56120.

XX

PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -

XX

PS Claim 11; SEQ ID NO 10932; 3071pp + Sequence Listing; English.

XX

CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.

XX

SQ Sequence 62 AA;

Query Match 73.5%; Score 36; DB 22; Length 62;

Best Local Similarity 100.0%; Pred. No. 56;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CLSSRLD 7

|||||

Db 53 CLSSRLD 59

RESULT 11

AAU29476

ID AAU29476 standard; Protein; 209 AA.

XX

AC AAU29476;

XX

DT 18-DEC-2001 (first entry)

XX

DE Human G protein-coupled receptor (GPCR) polypeptide #97.

XX

KW Human; G protein-coupled receptor; GPCR; mental disorder; schizophrenia;

KW neurological disorder; metabolic disorder; cancer; rheumatoid arthritis;

KW thyroid disorder; neurodegenerative disorder; cardiovascular disorder;

KW renal failure; autoimmune disorder; hyperproliferative disorder; HIV;

KW human immunodeficiency virus; viral infection; neuroprotective;

KW immunostimulant; neuroleptic; nootropic; tranquiliser; antidepressant;

KW anorectic; gene therapy.

XX

OS Homo sapiens.

XX

PN WO200168858-A2.

XX

PD 20-SEP-2001.

XX

PF 16-MAR-2001; 2001WO-US08456.

XX

PR 16-MAR-2000; 2000US-187783P.

PR 16-MAR-2000; 2000US-189907P.

PR 16-MAR-2000; 2000US-189917P.

PR 16-MAR-2000; 2000US-189918P.

PR 16-MAR-2000; 2000US-189960P.

PR 29-MAR-2000; 2000US-192155P.

PR 29-MAR-2000; 2000US-192234P.

PR 29-MAR-2000; 2000US-192830P.

PR 29-MAR-2000; 2000US-192916P.

PR 29-MAR-2000; 2000US-192923P.

PR 29-MAR-2000; 2000US-192933P.

PR 29-MAR-2000; 2000US-192945P.

XX

PA (PHAA) PHARMACIA & UPJOHN CO.

XX

PI Vogeli G;

XX

DR WPI; 2001-607458/69.

DR N-PSDB; AAS46915.

XX

PT Nucleic acid encoding G protein-coupled receptors, useful for the
PT prevention, diagnosis and treatment of mental disorders -

XX

PS Claim 31; Page 91; 274pp; English.

XX

CC Sequences AAU29380-AAU29509 represent human G protein-coupled receptor

CC (GPCR) polypeptides of the invention. The proteins and the DNA sequences

CC encoding them can be used to identify compounds which bind to GPCR

CC polypeptides and in screening for compounds that modulate GPCR activity.

CC By screening a human subject for the presence of mutations in GPCR DNA, a

CC GPCR-related disorder or a genetic predisposition can be diagnosed. The

CC sequences can also be used for treatment and prevention of mental
CC disorders such as schizophrenia, neurological disorders such as manic
CC depression, metabolic disorders such as obesity, cancer, rheumatoid
CC arthritis, thyroid disorders such as myxoedema, neurodegenerative
CC disorders such as Parkinson's disease, cardiovascular disorders such as
CC atherosclerosis, renal failure, autoimmune disorders, hyperproliferative
CC disorders such as psoriasis and viral infections such as those caused by
CC HIV.

XX

SQ Sequence 209 AA;

Query Match 73.5%; Score 36; DB 22; Length 209;
Best Local Similarity 55.6%; Pred. No. 1.8e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9

|:::||| |

Db 61 CMNNRLDPC 69

RESULT 12

ABG60764

ID ABG60764 standard; Protein; 209 AA.

XX

AC ABG60764;

XX

DT 13-AUG-2002 (first entry)

XX

DE Novel G protein coupled receptor (nGPCR-x) #97.

XX

KW G protein coupled receptor; nGPCR-x; immune response; thyroid disorder;
KW mental disorder; thyreotoxicosis; myxoedema; inflammatory condition;
KW Crohn's disease; cell differentiation; homeostasis; rheumatoid arthritis;
KW renal failure; autoimmune disorder; movement disorder; CNS disorder;
KW viral infection; human immunodeficiency virus; HIV; metabolic disorder;
KW cardiovascular disorder; diabetes; obesity; anorexia; cardiomyopathy;
KW porliferative disease; cancer; psoriasis; lung cancer; hormonal disorder;
KW sexual dysfunction.

XX

OS Homo sapiens.

XX

PN US2002058306-A1.

XX

PD 16-MAY-2002.

XX

PF 16-MAR-2001; 2001US-0811284.

XX

PR 16-MAR-2000; 2000US-189783P.

PR 16-MAR-2000; 2000US-189907P.

PR 16-MAR-2000; 2000US-189917P.

PR 16-MAR-2000; 2000US-189918P.

PR 16-MAR-2000; 2000US-189960P.

PR 24-MAR-2000; 2000US-192155P.

PR 27-MAR-2000; 2000US-192234P.

PR 29-MAR-2000; 2000US-192830P.

PR 29-MAR-2000; 2000US-192945P.

PR 29-MAR-2000; 2000US-192916P.

PR 29-MAR-2000; 2000US-192923P.
PR 29-MAR-2000; 2000US-192830P.
PR 29-MAR-2000; 2000US-192945P.
PR 29-MAR-2000; 2000US-192830P.
PR 29-MAR-2000; 2000US-192945P.
PR 29-MAR-2000; 2000US-192830P.
PR 29-MAR-2000; 2000US-192945P.

XX

PA (VOGE/) VOGELI G.

XX

PI Vogeli G;

XX

DR WPI; 2002-434856/46.

DR N-PSDB; ABK81693.

XX

PT New isolated nucleic acid encoding a G protein coupled receptor for
PT producing the receptor which can induce an immune response in a mammal
PT -

XX

PS Claim 27; Page 67; 216pp; English.

XX

CC The invention describes an isolated nucleic acid (I) comprising a
CC sequence encoding a portion of a G protein coupled receptor (nGPCR-x).
CC (I) is used to produce a recombinant nGPCR-x polypeptide. A polypeptide
CC encoded by (I) is used to induce an immune response in a mammal. nGPCR-x
CC is used to identify a compound that binds to it and/or modulates it's
CC activity. (I) is used to identify animal homologues of nGPCR-x. (I) can
CC be used to diagnose a human subject as having a brain or genetic
CC predisposition disorder, such as a mental disorder. (I) is used to
CC screen for an nGPCR-x related disorder including thyroid disorders (e.g.
CC thyreotoxicosis, myxoedema), renal failure, inflammatory conditions (e.g.
CC Crohn's disease), diseases related to cell differentiation and
CC homeostasis, rheumatoid arthritis, autoimmune disorders, movement
CC disorders, CNS disorders, viral infections (e.g. Human immunodeficiency
CC virus), metabolic and cardiovascular disorders (e.g. diabetes, obesity,
CC anorexia, cardiomyopathies), porliferative diseases and cancers (e.g.
CC psoriasis, lung cancer), hormonal disorders, sexual dysfunction and
CC hereditary mental disorders in a human patient. A host cell comprising
CC (I) is used to screen for a modulator of nGPCR-x activity. nGPCR-x is
CC used to identify compounds that can treat mental disorders. The
CC polypeptide encoded by (I) is used to purify a G protein from a sample.
CC This is the amino acid sequence of a novel G protein coupled receptor
CC (nGPCR-x) protein described in the invention.

XX

SQ Sequence 209 AA;

Query Match 73.5%; Score 36; DB 23; Length 209;
Best Local Similarity 55.6%; Pred. No. 1.8e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CLSSRLDAC 9
|::||| |
Db 61 CMNNRLDPC 69

RESULT 13
AAY01191

ID AAY01191 standard; Protein; 53 AA.
 XX
 AC AAY01191;
 XX
 DT 18-MAY-1999 (first entry)
 XX
 DE Polypeptide fragment encoded by gene 14.
 XX
 KW Human; secreted protein; gene therapy; protein therapy; tissue; cancer;
 KW tumour; neurodegenerative disorder; leukaemia; autoimmune disease; AIDS;
 KW developmental abnormality; foetal deficiency; Alzheimer's disease;
 KW cognitive disorder; schizophrenia; immunological disorder; mood disorder;
 KW immune deficiency disease; respiratory disorder; arthritis; skeletal;
 KW haematopoietic disorder; neural; osteoporosis; metabolic disorders;
 KW cardiovascular; endocrine; gastrointestinal; asthma; diagnosis.
 XX
 OS Homo sapiens.
 XX
 PN WO9901020-A2.
 XX
 PD 14-JAN-1999.
 XX
 PF 30-JUN-1998; 98WO-US13608.
 XX
 PR 12-SEP-1997; 97US-0058663.
 PR 01-JUL-1997; 97US-0051381.
 PR 01-JUL-1997; 97US-0051480.
 PR 12-SEP-1997; 97US-0058598.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Carter KC, Endress GA, Feng P, Rosen CA, Ruben SM;
 XX
 DR WPI; 1999-105683/09.
 DR N-PSDB; AAX22124.
 XX
 PT New isolated human genes and the secreted polypeptides they encode -
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, immune deficiency diseases or blood
 PT disorders
 XX
 PS Disclosure; Page 24; 179pp; English.
 XX
 CC The invention relates to nucleic acid sequences (AAX22111 to AAX22134)
 CC encoding human secreted proteins (AAY01135 to AAY01158). The secreted
 CC protein gene sequences are deposited with the ATCC under deposit number
 CC ATCC 209118. Host cells comprising recombinant vectors containing the
 CC nucleic acid sequences are used for the recombinant production of the
 CC secreted proteins. The polynucleotide and amino acid sequences are useful
 CC for are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Pathological conditions can
 CC be also diagnosed by determining the amount of the new polypeptides in a
 CC sample or by determining the presence of mutations in the new
 CC polynucleotides. Specific uses are described for each of the
 CC polynucleotides, based on which tissues they are most highly expressed
 CC in, and include developing products for the diagnosis or treatment of
 CC cancer, tumours, developmental abnormalities and foetal deficiencies,

CC autoimmune diseases, lymphomas, Alzheimer's and cognitive disorders,
CC schizophrenia, immunological disorders, immune deficiency diseases
CC (AIDS), mood disorders, respiratory disorders, arthritis, asthma,
CC haematopoietic disorders, neural disorders, skeletal disorders,
CC osteoporosis, metabolic disorders, cardiovascular disorders, endocrine
CC disorders or gastrointestinal disorders. The polypeptides are also useful
CC for identifying their binding partners. The present sequence represents a
CC polypeptide fragment encoded by a gene of the invention (see descriptor
CC line for gene number).
XX
SQ Sequence 53 AA;

Query Match 71.4%; Score 35; DB 20; Length 53;
Best Local Similarity 55.6%; Pred. No. 72;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
|:: ||| |
Db 40 CIAGRLDIC 48

RESULT 14
AAU61079

ID AAU61079 standard; Protein; 80 AA.
XX
AC AAU61079;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #21975.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59613.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for

PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1; SEQ ID No 22274; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent *Propionibacterium acnes* immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC *P. acnes*. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC *P. acnes* is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of *P. acnes* in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for *P. acnes* proteins. These antibodies can be used to
CC downregulate expression and activity of *P. acnes* polypeptides and
CC therefore treat *P. acnes* infections. The antibodies may also be used as
CC diagnostic agents for determining *P. acnes* presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 80 AA;

Query Match 71.4%; Score 35; DB 22; Length 80;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CLSSRLDA 8
||| |
Db 52 CLSSHLDA 59

RESULT 15

AAM06519

ID AAM06519 standard; Protein; 87 AA.

XX

AC AAM06519;

XX

DT 05-OCT-2001 (first entry)

XX

DE Human foetal protein, SEQ ID NO: 250.

XX

KW Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;
KW nootropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;
KW gene therapy; antisense therapy; cancer; immune disorder;
KW growth disorder; osteoporosis; thrombolytic disorder;
KW nervous system disorder; inflammation.

XX

OS Homo sapiens.

XX

PN WO200155339-A2.

XX

PD 02-AUG-2001.
 XX
 PF 25-JAN-2001; 2001WO-US02723.
 XX
 PR 25-JAN-2000; 2000US-0491404.
 PR 15-SEP-2000; 2000US-0663870.
 PR 06-NOV-2000; 2000US-0707351.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Yeung G, Ford JE, Boyle BJ, Arterburn MC, Drmanac RA, Tang YT;
 PI Liu C, Asundi V, Zhou P, Werhman T;
 XX
 DR WPI; 2001-465571/50.
 DR N-PSDB; AAH94194.
 XX
 PT Novel fetal proteins useful for the treatment and diagnosis of diseases
 PT associated with dysfunction of the protein e.g. cancers, immune
 PT disorders, growth disorders, thrombolytic disorders, nervous system
 PT disorders and inflammation -
 XX
 PS Claim 10; Page 271-272; 715pp; English.
 XX
 CC The invention relates to novel foetal polypeptides encoded by
 CC polynucleotides comprising one of 477 sequences fully defined in the
 CC specification. The foetal polynucleotides and polypeptides are
 CC useful in the treatment and diagnosis of diseases such as cancers,
 CC immune disorders, growth disorders (e.g. osteoporosis), thrombolytic
 CC disorders, nervous system disorders and inflammation. The present
 CC sequence is a polypeptide encoded by a cDNA assembled using
 CC an expressed sequence tag (EST) found to be expressed in human
 CC foetal tissue cDNA libraries.
 XX
 SQ Sequence 87 AA;

Query Match 71.4%; Score 35; DB 22; Length 87;
 Best Local Similarity 66.7%; Pred. No. 1.1e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CLSSRLDAC 9
 ||| ||::|
 Db 76 CLSVRLNSC 84

Search completed: November 13, 2003, 09:45:23
 Job time : 31.2812 secs

GenCore version 5.1.6
 Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:45:35 ; Search time 18.6562 Seconds
 (without alignments)
 88.069 Million cell updates/sec

Title: US-09-228-866-3
 Perfect score: 49
 Sequence: 1 CLSSRLDAC 9

 Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

 Searched: 666188 seqs, 182559486 residues

 Total number of hits satisfying chosen parameters: 666188

 Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

 Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : Published_Applications_AA:*
 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	% Query		DB	ID	Description
	Score	Match Length			
1	38	77.6	130	9 US-09-939-980-491	Sequence 491, App
2	38	77.6	297	15 US-10-156-761-13501	Sequence 13501, A
3	36	73.5	209	9 US-09-811-284-225	Sequence 225, App
4	34	69.4	114	9 US-09-820-893-65	Sequence 65, Appl
5	34	69.4	132	9 US-09-820-893-100	Sequence 100, App
6	34	69.4	1147	11 US-09-798-412-11	Sequence 11, Appl
7	34	69.4	1147	15 US-10-325-917-11	Sequence 11, Appl
8	34	69.4	1247	14 US-10-032-159A-8	Sequence 8, Appli
9	33	67.3	118	10 US-09-918-508-8	Sequence 8, Appli
10	33	67.3	175	10 US-09-764-864-877	Sequence 877, App

11	33	67.3	537	12	US-10-321-802-2	Sequence 2, Appli
12	33	67.3	617	11	US-09-866-050A-303	Sequence 303, App
13	32	65.3	27	15	US-10-203-860-6	Sequence 6, Appli
14	32	65.3	143	15	US-10-012-542-203	Sequence 203, App
15	32	65.3	180	15	US-10-275-555-2	Sequence 2, Appli
16	32	65.3	296	12	US-10-096-494-1	Sequence 1, Appli
17	32	65.3	319	12	US-10-017-161-1346	Sequence 1346, Ap
18	32	65.3	632	10	US-09-981-353-50	Sequence 50, Appl
19	31	63.3	91	10	US-09-738-626-4813	Sequence 4813, Ap
20	31	63.3	115	15	US-10-007-280A-205	Sequence 205, App
21	31	63.3	152	11	US-09-974-879-199	Sequence 199, App
22	31	63.3	153	11	US-09-305-736-199	Sequence 199, App
23	31	63.3	195	12	US-10-358-790-5	Sequence 5, Appli
24	31	63.3	211	15	US-10-193-460A-2	Sequence 2, Appli
25	31	63.3	212	12	US-10-029-386-34101	Sequence 34101, A
26	31	63.3	215	12	US-10-358-790-1	Sequence 1, Appli
27	31	63.3	220	15	US-10-193-460A-1	Sequence 1, Appli
28	31	63.3	222	15	US-10-208-948-10	Sequence 10, Appl
29	31	63.3	236	12	US-10-358-790-2	Sequence 2, Appli
30	31	63.3	238	10	US-09-738-626-5263	Sequence 5263, Ap
31	31	63.3	259	9	US-09-925-302-473	Sequence 473, App
32	31	63.3	439	9	US-09-159-469-60	Sequence 60, Appl
33	31	63.3	439	9	US-09-798-042-60	Sequence 60, Appl
34	31	63.3	491	9	US-09-815-242-10190	Sequence 10190, A
35	31	63.3	539	15	US-10-128-714-3459	Sequence 3459, Ap
36	31	63.3	564	15	US-10-234-432-29	Sequence 29, Appl
37	31	63.3	568	15	US-10-128-714-8459	Sequence 8459, Ap
38	31	63.3	897	15	US-10-043-487-287	Sequence 287, App
39	31	63.3	2871	12	US-10-015-115-57	Sequence 57, Appl
40	31	63.3	2871	12	US-10-015-115-59	Sequence 59, Appl
41	31	63.3	3002	12	US-10-015-115-56	Sequence 56, Appl
42	31	63.3	4545	9	US-09-873-403-2	Sequence 2, Appli
43	31	63.3	4861	10	US-09-919-497-70	Sequence 70, Appl
44	31	63.3	4861	15	US-10-097-534-26	Sequence 26, Appl
45	31	63.3	4861	15	US-10-146-473-49	Sequence 49, Appl

ALIGNMENTS

RESULT 1

US-09-939-980-491

; Sequence 491, Application US/09939980

; Patent No. US20020082234A1

; GENERAL INFORMATION:

; APPLICANT: Black, Michael
; Burnham, Martin
; Hodgson, John
; Knowles, David
; Lonetto, Michael
; Nicholas, Richard
; Pratt, Julie
; Reichard, Richard
; Rosenberg, Martin
; Ward, Judith

; TITLE OF INVENTION: No. US20020082234A1el Prokaryotic Polynucleotides,
; Polypeptides and Their Uses

```

;      NUMBER OF SEQUENCES: 534
;      CORRESPONDENCE ADDRESS:
;          ADDRESSEE: SmithKline Beecham Corporation
;          STREET: 709 Swedeland Road
;          CITY: King of Prussia
;          STATE: PA
;          COUNTRY: USA
;          ZIP: 19406-0939
;      COMPUTER READABLE FORM:
;          MEDIUM TYPE: Diskette
;          COMPUTER: IBM Compatible
;          OPERATING SYSTEM: DOS
;          SOFTWARE: FastSEQ for Windows Version 2.0
;      CURRENT APPLICATION DATA:
;          APPLICATION NUMBER: US/09/939,980
;          FILING DATE: 27-Aug-2001
;          CLASSIFICATION: <Unknown>
;      PRIOR APPLICATION DATA:
;          APPLICATION NUMBER: 08/936,165
;          FILING DATE: <Unknown>
;      ATTORNEY/AGENT INFORMATION:
;          NAME: Gimmi, Edward R
;          REGISTRATION NUMBER: 38,891
;          REFERENCE/DOCKET NUMBER: P50549
;      TELECOMMUNICATION INFORMATION:
;          TELEPHONE: 610-270-4478
;          TELEFAX: 610-270-5090
;          TELEX: <Unknown>
;      INFORMATION FOR SEQ ID NO: 491:
;          SEQUENCE CHARACTERISTICS:
;              LENGTH: 130 amino acids
;              TYPE: amino acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;          MOLECULE TYPE: Protein
;          SEQUENCE DESCRIPTION: SEQ ID NO: 491:
US-09-939-980-491

```

```

Query Match          77.6%;  Score 38;  DB 9;  Length 130;
Best Local Similarity 77.8%;  Pred. No. 13;
Matches      7;  Conservative    0;  Mismatches    2;  Indels      0;  Gaps      0;

```

```

Qy          1 CLSSRLDAC 9
             || || |||
Db          112 CLLSRCDAC 120

```

```

RESULT 2
US-10-156-761-13501
; Sequence 13501, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI

```

```

; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13501
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13501

```

```

Query Match          77.6%; Score 38; DB 15; Length 297;
Best Local Similarity 66.7%; Pred. No. 30;
Matches      6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 CLSSRLDAC 9
        || ::|||
Db      203 CLDAQLDAC 211

```

RESULT 3

```

US-09-811-284-225
; Sequence 225, Application US/09811284
; Patent No. US20020058306A1
; GENERAL INFORMATION:
; APPLICANT: Vogeli, Gabriel
; TITLE OF INVENTION: No. US20020058306A1el G Protein-Coupled Receptors
; FILE REFERENCE: 00167US1
; CURRENT APPLICATION NUMBER: US/09/811,284
; CURRENT FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/189,783
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/189,907
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/189,918
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/189,960
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/189,917
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/192,945
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/192,916
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/192,923
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/192,933
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/192,830
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/192,234

```

; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/192,155
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/192,935
; PRIOR FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 258
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 225
; LENGTH: 209
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-811-284-225

Query Match 73.5%; Score 36; DB 9; Length 209;
Best Local Similarity 55.6%; Pred. No. 49;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
|::||| |
Db 61 CMNNRLDPC 69

RESULT 4

US-09-820-893-65
; Sequence 65, Application US/09820893
; Patent No. US20020076705A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 31 Human Secreted Proteins
; FILE REFERENCE: PZ033P1
; CURRENT APPLICATION NUMBER: US/09/820,893
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/531,119
; PRIOR FILING DATE: 2000-03-20
; PRIOR APPLICATION NUMBER: 60/102,895
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 65
; LENGTH: 114
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (114)
; OTHER INFORMATION: Xaa equals stop translation
US-09-820-893-65

Query Match 69.4%; Score 34; DB 9; Length 114;
Best Local Similarity 55.6%; Pred. No. 64;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
|:| |:| |
Db 20 CISFRVDVC 28

RESULT 5

US-09-820-893-100

; Sequence 100, Application US/09820893
 ; Patent No. US20020076705A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: 31 Human Secreted Proteins
 ; FILE REFERENCE: PZ033P1
 ; CURRENT APPLICATION NUMBER: US/09/820,893
 ; CURRENT FILING DATE: 2001-03-30
 ; PRIOR APPLICATION NUMBER: 09/531,119
 ; PRIOR FILING DATE: 2000-03-20
 ; PRIOR APPLICATION NUMBER: 60/102,895
 ; PRIOR FILING DATE: 1998-10-02
 ; NUMBER OF SEQ ID NOS: 140
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 100
 ; LENGTH: 132
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-820-893-100

Query Match 69.4%; Score 34; DB 9; Length 132;
 Best Local Similarity 55.6%; Pred. No. 74;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
 |:| |:| |
 Db 39 CISFRVDVC 47

RESULT 6

US-09-798-412-11

; Sequence 11, Application US/09798412
 ; Publication No. US20030109428A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bertin, John
 ; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
 ; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
 ; FILE REFERENCE: 07334-327001
 ; CURRENT APPLICATION NUMBER: US/09/798,412
 ; CURRENT FILING DATE: 2001-03-02
 ; PRIOR APPLICATION NUMBER: US 09/728,260
 ; PRIOR FILING DATE: 2000-12-01
 ; PRIOR APPLICATION NUMBER: US 09/685,791
 ; PRIOR FILING DATE: 2000-10-10
 ; PRIOR APPLICATION NUMBER: US 09/513,904
 ; PRIOR FILING DATE: 2000-02-25
 ; PRIOR APPLICATION NUMBER: US 09/507,533
 ; PRIOR FILING DATE: 2000-02-18
 ; PRIOR APPLICATION NUMBER: US 60/168,780
 ; PRIOR FILING DATE: 1999-12-03
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 11
 ; LENGTH: 1147
 ; TYPE: PRT

; ORGANISM: Homo sapiens
US-09-798-412-11

Query Match 69.4%; Score 34; DB 11; Length 1147;
Best Local Similarity 75.0%; Pred. No. 5.8e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LSSRLDAC 9
: ||: ||||
Db 777 ISSQLDAC 784

RESULT 7

US-10-325-917-11

; Sequence 11, Application US/10325917
; Publication No. US20030113787A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-327001
; CURRENT APPLICATION NUMBER: US/10/325,917
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: US/09/798,412
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 09/728,260
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 09/685,791
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/513,904
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 09/507,533
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: US 60/168,780
; PRIOR FILING DATE: 1999-12-03
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1147
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-325-917-11

Query Match 69.4%; Score 34; DB 15; Length 1147;
Best Local Similarity 75.0%; Pred. No. 5.8e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LSSRLDAC 9
: ||: ||||
Db 777 ISSQLDAC 784

RESULT 8

US-10-032-159A-8

; Sequence 8, Application US/10032159A
; Publication No. US20020164703A1
; GENERAL INFORMATION:

```
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; TITLE OF INVENTION: CARD-DOMAIN CONTAINING POLYPEPTIDES,
; TITLE OF INVENTION: ENCODING NUCLEIC ACIDS, AND METHODS OF USE
; FILE REFERENCE: P-LJ 5100
; CURRENT APPLICATION NUMBER: US/10/032,159A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US 60/257,457
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1247
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032-159A-8
```

```
Query Match          69.4%; Score 34; DB 14; Length 1247;
Best Local Similarity 75.0%; Pred. No. 6.3e+02;
Matches      6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      2 LSSRLDAC 9
        :||:||||
Db      909 ISSQLDAC 916
```

RESULT 9

```
US-09-918-508-8
; Sequence 8, Application US/09918508
; Patent No. US20020177162A1
; GENERAL INFORMATION:
; APPLICANT: KAKIMOTO, TATSUO
; APPLICANT: HIGUCHI, MASAYUKI
; APPLICANT: INOUE, TSUTOMU
; TITLE OF INVENTION: ANALYSIS OF AGONIST-ACTIVITY AND ANTAGONIST-ACTIVITY
; TITLE OF INVENTION: TO CYTOKININ RECEPTOR
; FILE REFERENCE: Q65478
; CURRENT APPLICATION NUMBER: US/09/918,508
; CURRENT FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: JP 2001-073812
; PRIOR FILING DATE: 2001-03-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-918-508-8
```

```
Query Match          67.3%; Score 33; DB 10; Length 118;
Best Local Similarity 55.6%; Pred. No. 1e+02;
Matches      5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy      1 CLSSRLDAC 9
        || |:|
Db      93 CLESGMDSC 101
```


RESULT 10

US-09-764-864-877

; Sequence 877, Application US/09764864
 ; Patent No. US20020132753A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PTZ23
 ; CURRENT APPLICATION NUMBER: US/09/764,864
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 1792
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 877
 ; LENGTH: 175
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-764-864-877

Query Match 67.3%; Score 33; DB 10; Length 175;
 Best Local Similarity 66.7%; Pred. No. 1.5e+02;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
 |||||
 Db 5 CLSSRCSLC 13

RESULT 11

US-10-321-802-2

; Sequence 2, Application US/10321802
 ; Publication No. US20030200563A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Butler, Karlene H.
 ; APPLICANT: Cahoon, Edgar B.
 ; APPLICANT: Cahoon, Rebecca E.
 ; APPLICANT: Famodu, Omolayo O.
 ; APPLICANT: Hall, Sarah E.
 ; TITLE OF INVENTION: Phospholipid:diacylglycerol Acetyltransferases
 ; FILE REFERENCE: BB1486 US NA
 ; CURRENT APPLICATION NUMBER: US/10/321,802
 ; CURRENT FILING DATE: 2002-12-17
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 2
 ; LENGTH: 537
 ; TYPE: PRT
 ; ORGANISM: Momordica charantia
 US-10-321-802-2

Query Match 67.3%; Score 33; DB 12; Length 537;
 Best Local Similarity 66.7%; Pred. No. 4.3e+02;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9

Db

503 CPSSRAEAC 511

RESULT 12

US-09-866-050A-303

; Sequence 303, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 303
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Mouse
US-09-866-050A-303

Query Match 67.3%; Score 33; DB 11; Length 617;
Best Local Similarity 55.6%; Pred. No. 4.9e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy

1 CLSSRLDAC 9

|| :||: |

Db

584 CLKNRLEQC 592

RESULT 13

US-10-203-860-6

; Sequence 6, Application US/10203860
; Publication No. US20030108904A1
; GENERAL INFORMATION:
; APPLICANT: WAKAMIYA, No. US20030108904A1utaka
; TITLE OF INVENTION: No. US20030108904A1el Scavenger Receptor
; FILE REFERENCE: 19036/38693
; CURRENT APPLICATION NUMBER: US/10/203,860
; CURRENT FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: 2000-35155
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 2000-309068
; PRIOR FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 28
; SEQ ID NO 6
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:

; OTHER INFORMATION: Modified Consensus Sequence of collectins Hybridizable
with No. US20030108904A1e1
; OTHER INFORMATION: Collectin.
US-10-203-860-6

Query Match 65.3%; Score 32; DB 15; Length 27;
Best Local Similarity 66.7%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
||| ||| |
Db 17 CLQSRLAIC 25

RESULT 14

US-10-012-542-203
; Sequence 203, Application US/10012542
; Publication No. US20030044851A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 94 Human Secreted Proteins
; FILE REFERENCE: PZ029P1
; CURRENT APPLICATION NUMBER: US/10/012,542
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/461,325
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,507
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,508
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,509
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,510
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,112
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,113
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 203
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-542-203

Query Match 65.3%; Score 32; DB 15; Length 143;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 SRLDAC 9
|||||
Db 51 SRLDAC 56

RESULT 15

US-10-275-555-2

; Sequence 2, Application US/10275555
; Publication No. US20030104450A1
; GENERAL INFORMATION:
; APPLICANT: Merck Patent GmbH
; TITLE OF INVENTION: No. US20030104450A1 regulator of G protein signalling (RGS8)
; FILE REFERENCE: RGS8CWWS
; CURRENT APPLICATION NUMBER: US/10/275,555
; CURRENT FILING DATE: 2002-11-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-275-555-2

Query Match 65.3%; Score 32; DB 15; Length 180;
Best Local Similarity 55.6%; Pred. No. 2.3e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
||| : |:
Db 19 CLSHKSDSC 27

Search completed: November 13, 2003, 09:58:27
Job time : 18.6562 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:38:30 ; Search time 9.375 Seconds
(without alignments)
92.322 Million cell updates/sec

Title: US-09-228-866-3
Perfect score: 49
Sequence: 1 CLSSRLDAC 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*

1: pirl:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	41	83.7	2533	2	T28675	alpha-51D immobili
2	41	83.7	2533	2	T28674	alpha-51D-immobili
3	41	83.7	2543	2	T31687	surface antigen - P
4	37	75.5	92	2	G82701	hypothetical prote
5	37	75.5	152	2	F84115	hypothetical prote
6	37	75.5	383	1	S15624	E2 protein - human
7	35	71.4	153	2	G95290	hypothetical prote
8	35	71.4	161	2	T21405	hypothetical prote
9	35	71.4	255	2	S55641	uracil DNA glycosy
10	35	71.4	258	2	F72616	hypothetical prote
11	35	71.4	389	2	G87332	hypothetical prote
12	35	71.4	579	2	F69157	excinuclease ABC c
13	34	69.4	376	2	S36535	E2 protein - human
14	34	69.4	388	2	S36500	E2 protein - human
15	34	69.4	391	1	S15617	E2 protein - human
16	34	69.4	457	2	I51660	dopamine D1B recep
17	34	69.4	494	2	AB0252	glucose-6-phosphat
18	34	69.4	781	2	A56244	DNA repair/recombi
19	33	67.3	115	2	T36957	hypothetical prote
20	33	67.3	140	2	I54890	temperature-depend
21	33	67.3	280	2	H72593	hypothetical prote
22	33	67.3	305	2	S01875	hypothetical prote
23	33	67.3	306	2	T15858	hypothetical prote
24	33	67.3	394	2	S36512	E2 protein - human
25	33	67.3	398	1	W2WL42	E2 protein - human
26	33	67.3	613	2	T00853	hypothetical prote
27	33	67.3	615	2	JE0358	ralB binding prote
28	33	67.3	933	1	BVECCC	sensor protein rcs
29	33	67.3	933	2	C91017	sensor for ctr cap
30	33	67.3	933	2	E85861	hypothetical prote
31	33	67.3	948	2	AD0790	sensor protein Rcs
32	33	67.3	1199	2	T15826	hypothetical prote
33	32	65.3	75	2	JC2072	double-headed prot
34	32	65.3	76	2	JC2073	double-headed prot
35	32	65.3	213	2	E71511	hypothetical prote
36	32	65.3	317	2	T05528	hypothetical prote
37	32	65.3	369	2	I46531	surfactant protein
38	32	65.3	370	1	LNRRBB	pulmonary surfacta
39	32	65.3	419	2	H95919	probable reverse t
40	32	65.3	419	2	H96025	probable reverse t
41	32	65.3	419	2	A95386	Reverse transcript
42	32	65.3	567	2	T49942	hypothetical prote

43	32	65.3	621	2	JC7278	adaptor protein co
44	32	65.3	754	2	G90484	beta-xylosidase [i
45	32	65.3	852	2	S41886	DNA repair protein

ALIGNMENTS

RESULT 1

T28675

alpha-51D immobilization antigen - Paramecium tetraurelia

C;Species: Paramecium tetraurelia

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000

C;Accession: T28675

R;Schwegmann, K.J.

submitted to the EMBL Data Library, March 1996

A;Reference number: Z20506

A;Accession: T28675

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-2533 <SCH>

A;Cross-references: EMBL:X96400; PIDN:CAA65264.1

C;Genetics:

A;Gene: alpha-51D

A;Genetic code: SGC5

A;Introns: 280/3; 538/2; 1248/2

C;Superfamily: G surface protein

Query Match	83.7%;	Score 41;	DB 2;	Length 2533;
Best Local Similarity	66.7%;	Pred. No. 8.7;		
Matches	6;	Conservative	3;	Mismatches 0; Indels 0; Gaps 0;

Qy	1	CLSSRLDAC	9
		: : :	
Db	575	CISNRVDAC	583

RESULT 2

T28674

alpha-51D-immobilization antigen - Paramecium tetraurelia

C;Species: Paramecium tetraurelia

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000

C;Accession: T28674

R;Schmidt, H.J.

submitted to the EMBL Data Library, March 1995

A;Reference number: Z20505

A;Accession: T28674

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-2533 <SCH>

A;Cross-references: EMBL:X85135; NID:g728634; PID:g728635; PIDN:CAA59447.1

C;Genetics:

A;Genetic code: SGC5

A;Note: alpha-51D

C;Superfamily: G surface protein

Query Match	83.7%;	Score 41;	DB 2;	Length 2533;
-------------	--------	-----------	-------	--------------

Best Local Similarity 66.7%; Pred. No. 8.7;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
|:|:|:|
Db 575 CISNRVDAC 583

RESULT 3

T31687

surface antigen - *Paramecium primaurelia*

C;Species: *Paramecium primaurelia*

C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 23-Mar-2001

C;Accession: T31687

R;Bourgain-Guglielmetti, F.; Caron, F.

Journal of Eukaryot. Microbiol. 43, 303-314, 1996

A;Title: Molecular characterization of the D surface protein gene subfamily in *Paramecium primaurelia*.

A;Reference number: Z21061; MUID:96313351; PMID:8768434

A;Accession: T31687

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-2543 <BOU>

A;Cross-references: EMBL:X96616; NID:g1235576; PIDN:CAA65436.1

C;Genetics:

A;Genetic code: SGC5

C;Superfamily: G surface protein

Query Match 83.7%; Score 41; DB 2; Length 2543;
Best Local Similarity 66.7%; Pred. No. 8.7;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
|:|:|:|
Db 575 CISNRVDAC 583

RESULT 4

G82701

hypothetical protein XF1273 [imported] - *Xylella fastidiosa* (strain 9a5c)

C;Species: *Xylella fastidiosa*

C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C;Accession: G82701

R;anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis, Sao Paulo, Brazil.

Nature 406, 151-157, 2000

A;Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.

A;Reference number: A82515; MUID:20365717; PMID:10910347

A;Note: for a complete list of authors see reference number A59328 below

A;Accession: G82701

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-92 <SIM>

A;Cross-references: GB:AE003961; GB:AE003849; NID:g9106254; PIDN:AAF84082.1;

GSPDB:GN00128; XFSC:XF1273

A;Experimental source: strain 9a5c

R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Alves, L.M.C.; Araya, J.E.; Baia, G.S.; Baptista, C.S.; Barros, M.H.; Bonaccorsi, E.D.; Bordin, S.; Bove, J.M.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.; Colauto, N.B.; Colombo, C.; Costa, F.F.; Costa, M.C.R.; Costa-Neto, C.M.; Coutinho, L.L.; Cristofani, M.; Dias-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme, M.; Furlan, L.R.; Garnier, M.; Goldman, G.H.; Goldman, M.H.S.; Gomes, S.L.; Gruber, A.; Ho, P.L.; Hoheisel, J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigret, F.; Lambais, M.R.; Leite, L.C.C.; Lemos, E.G.M.; Lemos, M.V.F.; Lopes, S.A.; Lopes, C.R.; Machado, J.A.; Machado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.A.L.

A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; Monteiro-Vitorello, C.B.; Moon, D.H.; Nagai, M.A.; Nascimento, A.L.T.O.; Netto, L.E.S.; Nhani Jr., A.; Nobrega, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Paris, A.; Peixoto, B.R.; Pereira, G.A.G.; Pereira Jr., H.A.; Pesquero, J.B.; Quaggio, R.B.; Roberto, P.G.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki, H.E.

A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira, J.F.; Silvestri, M.L.Z.; Siqueira, W.J.; de Souza, A.A.; de Souza, A.P.; Terenzi, M.F.; Truffi, D.; Tsai, S.M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Zago, M.A.; Zatz, M.; Meidanis, J.; Setubal, J.C.

A;Reference number: A59328

A;Contents: annotation

C;Genetics:

A;Gene: XF1273

Query Match	75.5%;	Score 37;	DB 2;	Length 92;
Best Local Similarity	87.5%;	Pred. No. 2.9;		
Matches	7;	Conservative	1;	Mismatches 0; Indels 0; Gaps 0;

Qy	1	CLSSRLDA	8
		:	
Db	67	CLASRLDA	74

RESULT 5

F84115

hypothetical protein BH3726 [imported] - *Bacillus halodurans* (strain C-125)

C;Species: *Bacillus halodurans*

C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C;Accession: F84115

R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hiramata, C.; Nakamura, Y.; Ogasawara, N.; Kuhara, S.; Horikoshi, K.

Nucleic Acids Res. 28, 4317-4331, 2000

A;Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*.

A;Reference number: A83650; MUID:20512582; PMID:11058132

A;Accession: F84115

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-152 <STO>

A;Cross-references: GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BAB07445.1;
GSPDB:GN00137
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH3726

Query Match 75.5%; Score 37; DB 2; Length 152;
Best Local Similarity 77.8%; Pred. No. 4.6;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
|| ||| ||
Db 127 CLPSRLKAC 135

RESULT 6

S15624

E2 protein - human papillomavirus type 57

C;Species: human papillomavirus type 57

A;Note: host Homo sapiens (man)

C;Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 16-Jul-1999

C;Accession: S15624

R;Hirsch-Behnam, A.; Delius, H.; de Villiers, E.M.

Virus Res. 18, 81-98, 1990

A;Title: A comparative sequence analysis of two human papillomavirus (HPV) types 2a and 57.

A;Reference number: S15614; MUID:91188699; PMID:1964523

A;Accession: S15624

A;Molecule type: DNA

A;Residues: 1-383 <HIR>

A;Cross-references: EMBL:X55965; NID:g60882; PIDN:CAA39433.1; PID:g60886

C;Superfamily: papillomavirus E2 protein

C;Keywords: DNA binding; early protein; transcription regulation

Query Match 75.5%; Score 37; DB 1; Length 383;
Best Local Similarity 87.5%; Pred. No. 10;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LSSRLDAC 9
|:|||||
Db 4 LASRLDAC 11

RESULT 7

G95290

hypothetical protein Sma0443 [imported] - Sinorhizobium meliloti (strain 1021)
magaplasmid pSymA

C;Species: Sinorhizobium meliloti

C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C;Accession: G95290

R;Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bowser, L.; Capela, D.; Galibert, F.; Gouzy, J.; Gurjal, M.; Hong, A.; Huizar, L.; Hyman, R.W.; Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.; Davis, R.W.; Federspiel, N.A.; Long, S.R.

Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001

A;Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti pSymA megaplasmid.

A;Reference number: A95262; MUID:21396509; PMID:11481432

A;Accession: G95290

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-153 <KUR>

A;Cross-references: GB:AE006469; PIDN:AAK64889.1; PID:g14523307; GSPDB:GN00165

A;Experimental source: strain 1021, megaplasmid pSymA

R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, F.; Barnett, M.J.; Becker, A.; Boistard, P.; Bothe, G.; Boutry, M.; Bowser, L.; Buhrmester, J.; Cadieu, E.; Capela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; Gloux, S.; Godrie, T.; Goffeau, A.; Golding, B.; Gouzy, J.; Gurjal, M.; Hernandez-Lucas, I.; Hong, A.; Huizar, L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, V.; Masuy, D.; Palm, C.; Peck, M.C.; Pohl, T.M.; Portetelle, D.; Purnelle, B.; Ramsperger, U.; Surzycki, R.; Thebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.C.; Batut, J.

A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.

A;Reference number: A96039; MUID:21368234; PMID:11474104

A;Contents: annotation

C;Genetics:

A;Gene: Sma0443

A;Genome: plasmid

Query Match	71.4%;	Score 35;	DB 2;	Length 153;
Best Local Similarity	77.8%;	Pred. No. 11;		
Matches	7;	Conservative	0;	Mismatches 2; Indels 0; Gaps 0;

Qy	1	CLSSRLDAC	9
Db	138	CLPSRLMAC	146

RESULT 8

T21405

hypothetical protein F26D2.12 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C;Accession: T21405

R;McMurray, A.

submitted to the EMBL Data Library, November 1996

A;Reference number: Z19418

A;Accession: T21405

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-161 <WIL>

A;Cross-references: EMBL:Z81513; PIDN:CAB04182.1; GSPDB:GN00023; CESP:F26D2.12

A;Experimental source: clone F26D2

C;Genetics:

A;Gene: CESP:F26D2.12

A;Map position: 5

A;Introns: 24/1; 43/3; 79/1; 145/2

Query Match	71.4%;	Score 35;	DB 2;	Length 161;
-------------	--------	-----------	-------	-------------

Best Local Similarity 87.5%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLSSRLDA 8
||| |
Db 101 CLSSNLDA 108

RESULT 9

S55641

uracil DNA glycosylase - equine herpesvirus 2

C;Species: equine herpesvirus 2

C;Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 22-Jun-1999

C;Accession: S55641

R;Telford, E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.

J. Mol. Biol. 249, 520-528, 1995

A;Title: The DNA sequence of equine herpesvirus 2.

A;Reference number: S55594; MUID:95302501; PMID:7783207

A;Accession: S55641

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-255 <TEL>

A;Cross-references: GB:U20824; NID:g695172; PIDN:AAC13834.1; PID:g695219

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995

C;Superfamily: uracil-DNA glycosylase

Query Match 71.4%; Score 35; DB 2; Length 255;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
|||: |||
Db 176 CLSNELDHC 184

RESULT 10

F72616

hypothetical protein APE1391 - Aeropyrum pernix (strain K1)

C;Species: Aeropyrum pernix

C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C;Accession: F72616

R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.

DNA Res. 6, 83-101, 1999

A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix K1.

A;Reference number: A72450; MUID:99310339; PMID:10382966

A;Accession: F72616

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-258 <KAW>

A;Cross-references: DDBJ:AP000061; NID:g5104821; PIDN:BAA80388.1; PID:g5105074

A;Experimental source: strain K1

C;Genetics:

A;Gene: APE1391

C;Superfamily: Aeropyrum pernix hypothetical protein APE1391

Query Match 71.4%; Score 35; DB 2; Length 258;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
||| |||
Db 223 CLSGRLSTC 231

RESULT 11

G87332

hypothetical protein CC0674 [imported] - *Caulobacter crescentus*

C;Species: *Caulobacter crescentus*

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C;Accession: G87332

R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.F.; Alley, M.; Ohta, N.; Maddock, J.R.; Potocka, I.; Nelson, W.C.; Newton, A.; Stephens, C.; Phadke, N.D.; Ely, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonay, J.F.; Smit, J.; Craven, M.; Khouri, H.; Shetty, J.; Berry, K.; Utterback, T.; Tran, K.; Wolf, A.; Vamathevan, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of *Caulobacter crescentus*.

A;Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: G87332

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-389 <STO>

A;Cross-references: GB:AE005673; NID:g13421893; PIDN:AAK22659.1; GSPDB:GN00148

C;Genetics:

A;Gene: CC0674

Query Match 71.4%; Score 35; DB 2; Length 389;
Best Local Similarity 66.7%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
|| | |||
Db 44 CLPGRADAC 52

RESULT 12

F69157

excinuclease ABC chain C - *Methanobacterium thermoautotrophicum* (strain Delta H)

C;Species: *Methanobacterium thermoautotrophicum*

C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Apr-2000

C;Accession: F69157

R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso, A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shimer, G.; Goyal, A.;

Petrokovski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.;
Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta
H: functional analysis and comparative genomics.

A;Reference number: A69000; MUID:98037514; PMID:9371463

A;Accession: F69157

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-579 <MTH>

A;Cross-references: GB:AE000828; GB:AE000666; NID:g2621504; PIDN:AAB84947.1;
PID:g2621507

A;Experimental source: strain Delta H

C;Genetics:

A;Gene: MTH441

A;Start codon: TTG

C;Superfamily: excinuclease ABC chain C

Query Match 71.4%; Score 35; DB 2; Length 579;

Best Local Similarity 55.6%; Pred. No. 37;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
||:|::||
Db 156 CLNSQIDLC 164

RESULT 13

S36535

E2 protein - human papillomavirus type 10

C;Species: human papillomavirus type 10

C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-Aug-1999

C;Accession: S36535

R;Delius, H.; Hofmann, B.

submitted to the EMBL Data Library, August 1993

A;Description: Primer-directed sequencing of human papillomavirus types.

A;Reference number: S36469

A;Accession: S36535

A;Molecule type: DNA

A;Residues: 1-376

A;Cross-references: EMBL:X74465; NID:g396901; PIDN:CAA52492.1; PID:g396905

C;Superfamily: papillomavirus E2 protein

C;Keywords: DNA binding; early protein; transcription regulation

Query Match 69.4%; Score 34; DB 2; Length 376;

Best Local Similarity 75.0%; Pred. No. 40;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LSSRLDAC 9
|::|||||
Db 4 LANRLDAC 11

RESULT 14

S36500

E2 protein - human papillomavirus type 27

C;Species: human papillomavirus type 27

C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-Aug-1999
C;Accession: S36500
R;Delius, H.; Hofmann, B.
submitted to the EMBL Data Library, August 1993
A;Description: Primer-directed sequencing of human papillomavirus types.
A;Reference number: S36469
A;Accession: S36500
A;Molecule type: DNA
A;Residues: 1-388
A;Cross-references: EMBL:X74473; NID:g396964; PIDN:CAA52539.1; PID:g396968
C;Superfamily: papillomavirus E2 protein
C;Keywords: DNA binding; early protein; transcription regulation

Query Match 69.4%; Score 34; DB 2; Length 388;
Best Local Similarity 75.0%; Pred. No. 41;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LSSRLDAC 9
|:|||||
Db 4 LANRLDAC 11

RESULT 15

S15617

E2 protein - human papillomavirus type 2a

C;Species: human papillomavirus type 2a

A;Note: host Homo sapiens (man)

C;Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 16-Feb-1997

C;Accession: S15617

R;Hirsch-Behnam, A.; Delius, H.; de Villiers, E.M.

Virus Res. 18, 81-98, 1990

A;Title: A comparative sequence analysis of two human papillomavirus (HPV) types 2a and 57.

A;Reference number: S15614; MUID:91188699; PMID:1964523

A;Accession: S15617

A;Molecule type: DNA

A;Residues: 1-391 <HIR>

A;Cross-references: EMBL:X55964

C;Superfamily: papillomavirus E2 protein

C;Keywords: DNA binding; early protein; transcription regulation

Query Match 69.4%; Score 34; DB 1; Length 391;
Best Local Similarity 75.0%; Pred. No. 42;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LSSRLDAC 9
|:|||||
Db 4 LANRLDAC 11

Search completed: November 13, 2003, 09:52:52

Job time : 11.375 secs

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 5.15625 Seconds
(without alignments)
82.083 Million cell updates/sec

Title: US-09-228-866-3
Perfect score: 49
Sequence: 1 CLSSRLDAC 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	37	75.5	383	1	VE2_HPVS7	P22155 human papil
2	36	73.5	369	1	VE2_HPVS6	Q80958 human papil
3	35	71.4	255	1	UNG_HSVS2	P53765 equine herp
4	35	71.4	579	1	UVRC_METTH	O26541 methanobact
5	34	69.4	376	1	VE2_HPVS10	P36781 human papil
6	34	69.4	388	1	VE2_HPVS27	P36789 human papil
7	34	69.4	388	1	VE2_HPVS29	P50772 human papil
8	34	69.4	391	1	VE2_HPVS2A	P25482 human papil
9	34	69.4	457	1	DBDR_XENLA	P42290 xenopus lae
10	34	69.4	1147	1	CARB_HUMAN	Q9bx17 homo sapien
11	33	67.3	247	1	PSPA_CAVPO	P50403 cavia porce
12	33	67.3	305	1	GP7D_CHLTR	P10561 chlamydia t
13	33	67.3	382	1	VE2_HPVS61	Q80951 human papil
14	33	67.3	394	1	VE2_HPVS32	P36791 human papil
15	33	67.3	398	1	VE2_HPVS42	P27223 human papil
16	33	67.3	449	1	EF1C_PORPU	P50256 porphyra pu
17	33	67.3	948	1	RCSC_SALTI	Q56128 salmonella
18	33	67.3	948	1	RCSC_SALTY	P58662 salmonella
19	33	67.3	949	1	RCSC_ECOLI	P14376 escherichia
20	32	65.3	180	1	RGS8_HUMAN	P57771 homo sapien
21	32	65.3	180	1	RGS8_RAT	P49804 rattus norv
22	32	65.3	370	1	PSPB_RABIT	P15285 oryctolagus

23	32	65.3	852	1	RA54_SCHPO	P41410	schizosacch
24	31.5	64.3	134	1	FOLB_CHLPN	Q9z7e9	chlamydia p
25	31	63.3	78	1	IBB2_PHAAN	P01061	phaseolus a
26	31	63.3	83	1	IBB_PHALU	P01056	phaseolus l
27	31	63.3	111	1	UL91_HCMVA	P16797	human cytom
28	31	63.3	248	1	PSPA_HUMAN	P07714	homo sapien
29	31	63.3	367	1	VE2_HPVI1	P04015	human papil
30	31	63.3	368	1	VE2_HPVI6A	Q84294	human papil
31	31	63.3	368	1	VE2_HPVI6B	P03119	human papil
32	31	63.3	378	1	VE2_HPVI30	P36790	human papil
33	31	63.3	384	1	VE2_HPVI53	P36797	human papil
34	31	63.3	391	1	PCL_ECTHA	P42516	ectothiorho
35	31	63.3	398	1	DXR_YERPE	Q8zh62	yersinia pe
36	31	63.3	401	1	DXR_VIBPA	Q87me3	vibrio para
37	31	63.3	402	1	DXR_VIBCH	Q9kpv8	vibrio chol
38	31	63.3	433	1	THIC_FUSNN	Q8ri60	fusobacteri
39	31	63.3	491	1	G6PD_ECOLI	P22992	escherichia
40	31	63.3	491	1	G6PD_ERWCH	P37986	erwinia chr
41	31	63.3	498	1	KPYK_TRYBO	Q27788	trypanoplas
42	31	63.3	602	1	YHOH_SCHPO	O94364	schizosacch
43	31	63.3	702	1	ATI1_VARV	P34011	variola vir
44	31	63.3	724	1	ATI1_VACCV	P24759	vaccinia vi
45	31	63.3	726	1	ATI_CAMPC	Q05482	camelpox vi

ALIGNMENTS

RESULT 1

VE2_HPVI57

ID VE2_HPVI57 STANDARD; PRT; 383 AA.

AC P22155;

DT 01-AUG-1991 (Rel. 19, Created)

DT 01-AUG-1991 (Rel. 19, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE Regulatory protein E2.

GN E2.

OS Human papillomavirus type 57.

OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;

OC Papillomavirus.

OX NCBI_TaxID=10597;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=91188699; PubMed=1964523;

RA Hirsch-Behnam A., Delius H., de Villiers E.M.;

RT "A comparative sequence analysis of two human papillomavirus (HPV) types 2a and 57.";

RL Virus Res. 18:81-98(1990).

CC -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.

CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACCNNNNNNGGT-3') PRESENT

CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER

CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION

CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS

CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION

CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA

CC REPLICATION.

CC -!- SUBUNIT: Binds DNA as a dimer.

CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X55965; CAA39433.1; -.
 DR PIR; S15624; S15624.
 DR HSSP; P17383; 1DHM.
 DR InterPro; IPR000427; E2_C.
 DR InterPro; IPR001866; E2_N.
 DR Pfam; PF00511; E2_C; 1.
 DR Pfam; PF00508; E2_N; 1.
 DR ProDom; PD000672; E2_C; 1.
 DR ProDom; PD000678; E2_N; 1.
 KW Early protein; Transcription regulation; Activator; DNA-binding;
 KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
 SQ SEQUENCE 383 AA; 42829 MW; 7F20146677D7AAEC CRC64;

Query Match 75.5%; Score 37; DB 1; Length 383;
 Best Local Similarity 87.5%; Pred. No. 4.2;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LSSRLDAC 9
 | : | | | | | |
 Db 4 LASRLDAC 11

RESULT 2

VE2_HP66
 ID VE2_HP66 STANDARD; PRT; 369 AA.
 AC Q80958;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Regulatory protein E2.
 GN E2.
 OS Human papillomavirus type 66.
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
 OC Papillomavirus.
 OX NCBI_TaxID=37119;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Delius H.;
 RL Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.

CC -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
 CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACCNNNNNNGGT-3') PRESENT
 CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
 CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
 CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
 CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
 CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
 CC REPLICATION.

CC -!- SUBUNIT: Binds DNA as a dimer.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U31794; AAA79502.1; -.
DR HSSP; P17383; 1DHM.
DR InterPro; IPR000427; E2_C.
DR InterPro; IPR001866; E2_N.
DR Pfam; PF00511; E2_C; 1.
DR Pfam; PF00508; E2_N; 1.
DR ProDom; PD000672; E2_C; 1.
DR ProDom; PD000678; E2_N; 1.
KW Early protein; Transcription regulation; Activator; DNA-binding;
KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
SQ SEQUENCE 369 AA; 42781 MW; E90F265AEC397A14 CRC64;

Query Match 73.5%; Score 36; DB 1; Length 369;
Best Local Similarity 87.5%; Pred. No. 6.4;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LSSRLDAC 9
 || |||||
Db 4 LSQRLDAC 11

RESULT 3

UNG_HSVE2
ID UNG_HSVE2 STANDARD; PRT; 255 AA.
AC P53765;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Uracil-DNA glycosylase (EC 3.2.2.-) (UDG).
GN 46.
OS Equine herpesvirus type 2 (strain 86/87) (EHV-2).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae.
OX NCBI_TaxID=82831;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95302501; PubMed=7783207;
RA Telford E.A., Watson M.S., Aird H.C., Perry J., Davison A.J.;
RT "The DNA sequence of equine herpesvirus 2."
RL J. Mol. Biol. 249:520-528(1995).
CC -!- FUNCTION: EXCISES URACIL RESIDUES FROM THE DNA WHICH CAN ARISE
CC AS A RESULT OF MISINCORPORATION OF DUMP RESIDUES BY DNA
CC POLYMERASE OR DUE TO DEAMINATION OF CYTOSINE.
CC -!- SIMILARITY: BELONGS TO THE URACIL-DNA GLYCOSYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; U20824; AAC13834.1; -.
DR PIR; S55641; S55641.
DR HSSP; P12295; 3EUG.
DR InterPro; IPR003249; U_glycsylse_notp.
DR InterPro; IPR002043; UDNA_glycsylse.
DR InterPro; IPR005122; UDNA_glycsylseSF.
DR Pfam; PF03167; UDG; 1.
DR ProDom; PD001589; U_glycsylse_notp; 1.
DR TIGRFAMs; TIGR00628; ung; 1.
DR PROSITE; PS00130; U_DNA_GLYCOSYLASE; 1.
KW DNA repair; Hydrolase; Glycosidase.
FT ACT_SITE 90 90 GENERAL BASE (BY SIMILARITY).
SQ SEQUENCE 255 AA; 29099 MW; 20104402C5297336 CRC64;

Query Match 71.4%; Score 35; DB 1; Length 255;
Best Local Similarity 66.7%; Pred. No. 6.9;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
|||: |||
Db 176 CLSNELDHC 184

RESULT 4

UVRC_METTH

ID UVRC_METTH STANDARD; PRT; 579 AA.
AC O26541;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE UvrABC system protein C (UvrC protein) (Excinuclease ABC subunit C).
GN UVRC OR MTH441.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics."
RL J. Bacteriol. 179:7135-7155(1997).

```

CC  -!- FUNCTION: The UvrABC repair system catalyzes the recognition and
CC      processing of DNA lesions. UvrC both incises the 5' and 3' sides
CC      of the lesion. The N-terminal half is responsible for the 3'
CC      incision and the C-terminal half is responsible for the 5'
CC      incision (By similarity).
CC  -!- SUBUNIT: Interacts with uvrB in an incision complex (By
CC      similarity).
CC  -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC  -!- SIMILARITY: Belongs to the uvrC family.
CC  -!- SIMILARITY: Contains 1 UVR domain.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR    EMBL; AE000828; AAB84947.1; -.
DR    PIR; F69157; F69157.
DR    HSSP; P07025; 1E52.
DR    HAMAP; MF_00203; -; 1.
DR    InterPro; IPR000445; HhH.
DR    InterPro; IPR003583; HHH_1.
DR    InterPro; IPR001943; UvrB/C.
DR    InterPro; IPR004791; UvrC.
DR    InterPro; IPR001162; UvrC_C.
DR    InterPro; IPR000305; UvrC_N.
DR    Pfam; PF01541; Exci_endo_N; 1.
DR    Pfam; PF00633; HHH; 2.
DR    Pfam; PF02151; UVR; 1.
DR    ProDom; PD005870; UvrC_C; 1.
DR    SMART; SM00465; GIYc; 1.
DR    SMART; SM00278; HhH1; 2.
DR    TIGRFAMs; TIGR00194; uvrC; 1.
DR    PROSITE; PS50151; UVR; 1.
DR    PROSITE; PS50164; UVRC_1; 1.
DR    PROSITE; PS50165; UVRC_2; 1.
KW    SOS response; Excision nuclease; DNA repair; DNA recombination;
KW    DNA excision; Complete proteome.
FT    DOMAIN      193      228      UVR.
SQ    SEQUENCE    579 AA;  66293 MW;  83D3DF7B8F9E3A68 CRC64;

```

```

Query Match          71.4%;  Score 35;  DB 1;  Length 579;
Best Local Similarity 55.6%;  Pred. No. 16;
Matches      5;  Conservative      3;  Mismatches      1;  Indels      0;  Gaps      0;

```

```

Qy      1 CLSSRLDAC 9
        ||::||
Db      156 CLNSQIDLC 164

```

```

RESULT 5
VE2_HPVI0
ID  VE2_HPVI0      STANDARD;      PRT;      376 AA.
AC  P36781;

```

DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Regulatory protein E2.
 GN E2.
 OS Human papillomavirus type 10.
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
 OC Papillomavirus.
 OX NCBI_TaxID=10603;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94265501; PubMed=8205838;
 RA Delius H., Hofmann B.;
 RT "Primer-directed sequencing of human papillomavirus types.";
 RL Curr. Top. Microbiol. Immunol. 186:13-31(1994).
 CC -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
 CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACCNNNNNNGGT-3') PRESENT
 CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
 CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
 CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
 CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
 CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
 CC REPLICATION.
 CC -!- SUBUNIT: Binds DNA as a dimer.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X74465; CAA52492.1; -.
 DR PIR; S36535; S36535.
 DR HSSP; P17383; 1DHM.
 DR InterPro; IPR000427; E2_C.
 DR InterPro; IPR001866; E2_N.
 DR Pfam; PF00511; E2_C; 1.
 DR Pfam; PF00508; E2_N; 1.
 DR ProDom; PD000672; E2_C; 1.
 DR ProDom; PD000678; E2_N; 1.
 KW Early protein; Transcription regulation; Activator; DNA-binding;
 KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
 SQ SEQUENCE 376 AA; 43003 MW; 916B14B7FC51D7D1 CRC64;

Query Match 69.4%; Score 34; DB 1; Length 376;
 Best Local Similarity 75.0%; Pred. No. 17;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LSSRLDAC 9
 |::|||||
 Db 4 LANRLDAC 11

VE2_HPV27
ID VE2_HPV27 STANDARD; PRT; 388 AA.
AC P36789;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Regulatory protein E2.
GN E2.
OS Human papillomavirus type 27.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=31550;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94265501; PubMed=8205838;
RA Delius H., Hofmann B.;
RT "Primer-directed sequencing of human papillomavirus types.";
RL Curr. Top. Microbiol. Immunol. 186:13-31(1994).
CC -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACCNNNNNNGGT-3') PRESENT
CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC REPLICATION.
CC -!- SUBUNIT: Binds DNA as a dimer.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X74473; CAA52539.1; -.
DR PIR; S36500; S36500.
DR HSSP; P17383; 1DHM.
DR InterPro; IPR000427; E2_C.
DR InterPro; IPR001866; E2_N.
DR Pfam; PF00511; E2_C; 1.
DR Pfam; PF00508; E2_N; 1.
DR ProDom; PD000672; E2_C; 1.
DR ProDom; PD000678; E2_N; 1.
KW Early protein; Transcription regulation; Activator; DNA-binding;
KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
SQ SEQUENCE 388 AA; 43297 MW; 1C2740BA5C2C873B CRC64;

Query Match 69.4%; Score 34; DB 1; Length 388;
Best Local Similarity 75.0%; Pred. No. 17;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LSSRLDAC 9
|::|||||
Db 4 LANRLDAC 11

RESULT 7

VE2_HPV29

ID VE2_HPV29 STANDARD; PRT; 388 AA.
 AC P50772;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Regulatory protein E2.
 GN E2.
 OS Human papillomavirus type 29.
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
 OC Papillomavirus.
 OX NCBI_TaxID=37112;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Delius H.;
 RL Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U31784; AAA79432.1; -.
 DR HSSP; P17383; 1DHM.
 DR InterPro; IPR000427; E2_C.
 DR InterPro; IPR001866; E2_N.
 DR Pfam; PF00511; E2_C; 1.
 DR Pfam; PF00508; E2_N; 1.
 DR ProDom; PD000672; E2_C; 1.
 DR ProDom; PD000678; E2_N; 1.
 SQ SEQUENCE 388 AA; 44332 MW; 54422F4CD0613692 CRC64;

Query Match 69.4%; Score 34; DB 1; Length 388;
 Best Local Similarity 75.0%; Pred. No. 17;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LSSRLDAC 9
 |::|
 Db 4 LANRLDAC 11

RESULT 8

VE2_HPV2A

ID VE2_HPV2A STANDARD; PRT; 391 AA.
 AC P25482;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Regulatory protein E2.
 GN E2.
 OS Human papillomavirus type 2a.

OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
 OC Papillomavirus.
 OX NCBI_TaxID=10584;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91188699; PubMed=1964523;
 RA Hirsch-Behnam A., Delius H., de Villiers E.M.;
 RT "A comparative sequence analysis of two human papillomavirus (HPV)
 RT types 2a and 57.";
 RL Virus Res. 18:81-98(1990).
 CC -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
 CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACCNNNNNNGGT-3') PRESENT
 CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
 CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
 CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
 CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
 CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
 CC REPLICATION.
 CC -!- SUBUNIT: Binds DNA as a dimer.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X55964; -; NOT_ANNOTATED_CDS.
 DR PIR; S15617; S15617.
 DR HSSP; P17383; 1DHM.
 DR InterPro; IPR000427; E2_C.
 DR InterPro; IPR001866; E2_N.
 DR Pfam; PF00511; E2_C; 1.
 DR Pfam; PF00508; E2_N; 1.
 DR ProDom; PD000672; E2_C; 1.
 DR ProDom; PD000678; E2_N; 1.
 KW Early protein; Transcription regulation; Activator; DNA-binding;
 KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
 SQ SEQUENCE 391 AA; 43233 MW; 6F3862CD4A124B58 CRC64;

Query Match 69.4%; Score 34; DB 1; Length 391;
 Best Local Similarity 75.0%; Pred. No. 17;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LSSRLDAC 9
 |::|
 Db 4 LANRLDAC 11

RESULT 9
 DBDR_XENLA
 ID DBDR_XENLA STANDARD; PRT; 457 AA.
 AC P42290;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE D(1B) dopamine receptor (D(5) dopamine receptor).
 OS *Xenopus laevis* (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 OC Xenopodinae; *Xenopus*.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95024150; PubMed=7937989;
 RA Sugamori K.S., Demchyshyn L.L., Chung M., Niznik H.B.;
 RT "D1A, D1B, and D1C dopamine receptors from *Xenopus laevis*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:10536-10540(1994).
 CC -!- FUNCTION: THIS IS ONE OF THE FIVE TYPES (D1 TO D5) OF RECEPTORS
 CC FOR DOPAMINE. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G
 CC PROTEINS WHICH ACTIVATE ADENYLYL CYCLASE.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: BRAIN AND KIDNEY.
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U07864; AAA50829.1; -.
 DR PIR; I51660; I51660.
 DR InterPro; IPR000276; GPCR_Rhodpsn.
 DR Pfam; PF00001; 7tm_1; 1.
 DR PRINTS; PR00237; GPCRRHODOPSN.
 DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
 DR PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein;
 KW Multigene family; Lipoprotein; Palmitate.
 FT DOMAIN 1 41 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 42 67 1 (POTENTIAL).
 FT DOMAIN 68 78 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 79 105 2 (POTENTIAL).
 FT DOMAIN 106 114 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 115 137 3 (POTENTIAL).
 FT DOMAIN 138 156 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 157 181 4 (POTENTIAL).
 FT DOMAIN 182 205 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 206 231 5 (POTENTIAL).
 FT DOMAIN 232 282 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 283 309 6 (POTENTIAL).
 FT DOMAIN 310 326 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 327 351 7 (POTENTIAL).
 FT DOMAIN 352 457 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 114 199 BY SIMILARITY.
 FT LIPID 361 361 PALMITATE (BY SIMILARITY).
 SQ SEQUENCE 457 AA; 51656 MW; A0A389311E4CD2FB CRC64;

Query Match 69.4%; Score 34; DB 1; Length 457;
Best Local Similarity 55.6%; Pred. No. 20;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
| | : | : | : |
Db 254 CRSNRVDSC 262

RESULT 10

CARB_HUMAN

ID CARB_HUMAN STANDARD; PRT; 1147 AA.
AC Q9BXL7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Caspase recruitment domain protein 11 (CARD-containing MAGUK protein
DE 3) (Carma 1).
GN CARD11 OR CARMA1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21192234; PubMed=11278692;
RA Bertin J., Wang L., Guo Y., Jacobson M.D., Poyet J.-L.,
RA Srinivasula S.M., Merriam S., DiStefano P.S., Alnemri E.S.;
RT "CARD11 and CARD14 are novel caspase recruitment domain
RT (CARD)/membrane-associated guanylate kinase (MAGUK) family members
RT that interact with Bcl10 and activate NF-kappaB.";
RL J. Biol. Chem. 276:11877-11882(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21255663; PubMed=11356195;
RA Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschopp J.;
RT "Carma1, a CARD-containing binding partner of Bcl10, induces Bcl10
RT phosphorylation and NF-kappaB activation.";
RL FEBS Lett. 496:121-127(2001).
RN [3]
RP ERRATUM.
RA Gaide O., Martinon F., Micheau O., Bonnet D., Thome M., Tschopp J.;
RL FEBS Lett. 505:198-198(2001).
CC -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK. Stimulates the
CC phosphorylation of Bcl10.
CC -!- SUBUNIT: CARD11 and Bcl10 bind to each other by CARD-CARD
CC interaction.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: Detected in adult peripheral blood leukocytes,
CC thymus, spleen and liver. Also found in promyelocytic leukemia HL-
CC 60 cells, chronic myelogenous leukemia K562 cells, Burkitt's
CC lymphoma Raji cells and colorectal adenocarcinoma SW480 cells. Not
CC detected in HeLa S3, Molt-4, A549 and G431 cells.
CC -!- SIMILARITY: Contains 1 CARD domain.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -!- SIMILARITY: Contains 1 guanylate kinase-like domain.
CC -!- CAUTION: Supposed to contain a SH3 domain which is not detected by

CC PROSITE, Pfam or SMART.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF322641; AAG53402.1; -.

DR Genew; HGNC:16393; CARD11.

DR MIM; 607210; -.

DR GO; GO:0005624; C:membrane fraction; NAS.

DR GO; GO:0004384; F:membrane-associated guanylate kinase; NAS.

DR GO; GO:0005515; F:protein binding activity; IPI.

DR GO; GO:0007250; P:activation of NF-kappaB-inducing kinase; NAS.

DR InterPro; IPR001315; CARD.

DR InterPro; IPR000619; Guanylate_kin.

DR InterPro; IPR001478; PDZ.

DR SMART; SM00228; PDZ; 1.

DR PROSITE; PS50209; CARD; 1.

DR PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.

DR PROSITE; PS50052; GUANYLATE_KINASE_2; FALSE_NEG.

DR PROSITE; PS50106; PDZ; FALSE_NEG.

KW Coiled coil.

FT DOMAIN 11 103 CARD.

FT DOMAIN 123 442 COILED COIL (POTENTIAL).

FT DOMAIN 673 748 PDZ.

FT DOMAIN 966 1133 GUANYLATE KINASE.

FT CONFLICT 808 808 P -> L (IN REF. 2).

SQ SEQUENCE 1147 AA; 132641 MW; 913A4B015D2B36CC CRC64;

Query Match 69.4%; Score 34; DB 1; Length 1147;

Best Local Similarity 75.0%; Pred. No. 53;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LSSRLDAC 9

:||:||||

Db 777 ISSQLDAC 784

RESULT 11

PSPA_CAVPO

ID PSPA_CAVPO STANDARD; PRT; 247 AA.

AC P50403;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A)

DE (PSAP).

GN SFTPA1 OR SFTPA OR SFTP1.

OS Cavia porcellus (Guinea pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.

OX NCBI_TaxID=10141;

RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=Hartley; TISSUE=Lung;
 RX MEDLINE=98018900; PubMed=9357868;
 RA Yuan H.T., Gowan S., Kelly F.J., Bingle C.D.;
 RT "Cloning of guinea pig surfactant protein A defines a distinct
 RT cellular distribution pattern within the lung."
 RL Am. J. Physiol. 273:L900-L906(1997).
 CC -!- FUNCTION: IN PRESENCE OF CALCIUM IONS, PSAP BINDS TO SURFACTANT
 CC PHOSPHOLIPIDS AND CONTRIBUTES TO LOWER THE SURFACE TENSION AT THE
 CC AIR-LIQUID INTERFACE IN THE ALVEOLI OF THE MAMMALIAN LUNG AND IS
 CC ESSENTIAL FOR NORMAL RESPIRATION.
 CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.
 CC -!- SUBCELLULAR LOCATION: Extracellular.
 CC -!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
 CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
 CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
 CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
 CC -!- SIMILARITY: Contains 1 collagenous domain.
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U40869; AAB82952.1; -.
 DR HSSP; P22897; 1EGG.
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR001304; Lectin_C.
 DR Pfam; PF01391; Collagen; 1.
 DR Pfam; PF00059; lectin_c; 1.
 DR ProDom; PD000007; Clg_helix; 1.
 DR SMART; SM00034; CLECT; 1.
 DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
 DR PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
 KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
 KW Signal; Lectin; Collagen; Repeat.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 247 PULMONARY SURFACTANT-ASSOCIATED PROTEIN
 FT A.
 FT DOMAIN 27 99 COLLAGEN-LIKE.
 FT DOMAIN 152 245 C-TYPE LECTIN (SHORT FORM).
 FT DISULFID 154 245 BY SIMILARITY.
 FT DISULFID 223 237 BY SIMILARITY.
 FT CARBOHYD 20 20 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 206 206 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 247 AA; 26104 MW; D1BC86270EEFC932 CRC64;

Query Match 67.3%; Score 33; DB 1; Length 247;
 Best Local Similarity 66.7%; Pred. No. 17;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
 || ||| |

RESULT 12

GP7D_CHLTR

ID GP7D_CHLTR STANDARD; PRT; 305 AA.
AC P10561; Q46427;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Virulence plasmid integrase pGP7-D (Protein P-11).
OS Chlamydia trachomatis.
OG Plasmid pLGV440, Plasmid pCHL1, and Plasmid pCTT1.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L2/434/Bu; PLASMID=pLGV440;
RX MEDLINE=89013895; PubMed=2845228;
RA Comanducci M., Ricci S., Ratti G.;
RT "The structure of a plasmid of Chlamydia trachomatis believed to be
RT required for growth within mammalian cells.";
RL Mol. Microbiol. 2:531-538(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=L1/440/LN; PLASMID=pLGV440;
RX MEDLINE=88233998; PubMed=2836808;
RA Hatt C., Ward M.E., Clarke I.N.;
RT "Analysis of the entire nucleotide sequence of the cryptic plasmid of
RT Chlamydia trachomatis serovar L1. Evidence for involvement in DNA
RT replication.";
RL Nucleic Acids Res. 16:4053-4067(1988).
RN [3]
RP REVISIONS.
RC STRAIN=L1/440/LN; PLASMID=pLGV440;
RA Hatt C.;
RL Submitted (APR-1994) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=G0/86 / Serotype D; PLASMID=pCHL1;
RX MEDLINE=90301796; PubMed=2194229;
RA Comanducci M., Ricci S., Cevenini R., Ratti G.;
RT "Diversity of the Chlamydia trachomatis common plasmid in biovars
RT with different pathogenicity.";
RL Plasmid 23:149-154(1990).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=Serotype B; PLASMID=pCTT1;
RX MEDLINE=88177106; PubMed=3444859;
RA Sriprakash K.S., Macavoy E.S.;
RT "Characterization and sequence of a plasmid from the trachoma biovar
RT of Chlamydia trachomatis.";
RL Plasmid 18:205-214(1987).
CC -!- MISCELLANEOUS: PGP7-D IS REQUIRED FOR GROWTH WITHIN MAMMALIAN
CC CELLS.
CC -!- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF PLASMID PLGV440.
CC -!- SIMILARITY: BELONGS TO THE "PHAGE" INTEGRASE FAMILY.

CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT.
CC -!- CAUTION: REF.5 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 254.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X07547; CAA30427.1; ALT_FRAME.
DR EMBL; X06707; CAA29890.1; -.
DR EMBL; J03321; AAA91567.1; -.
DR EMBL; M19487; AAB02592.1; ALT_INIT.
DR EMBL; M19487; AAB02584.1; ALT_FRAME.
DR PIR; S01875; S01875.
DR InterPro; IPR002104; Phage_integrase.
DR Pfam; PF00589; Phage_integrase; 1.
KW DNA recombination; DNA integration; Plasmid.
FT ACT_SITE 289 289 TRANSIENT COVALENT LINKAGE TO DNA DURING
FT STRAND CLEAVAGE AND REJOINING (BY
FT SIMILARITY).
FT VARIANT 28 28 H -> Y (IN PLASMID PCHL1).
FT VARIANT 244 244 Y -> H (IN PLASMIDS PCHL1 AND PCTT1).
FT VARIANT 296 296 S -> I (IN PLASMID PCHL1).
FT VARIANT 303 303 P -> T (IN PLASMIDS PCHL1 AND PCTT1).
SQ SEQUENCE 305 AA; 34805 MW; 048C77FB84A42C19 CRC64;

Query Match 67.3%; Score 33; DB 1; Length 305;
Best Local Similarity 66.7%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
 |||||
Db 279 CLSSRQSV 287

RESULT 13
VE2_HPV61
ID VE2_HPV61 STANDARD; PRT; 382 AA.
AC Q80951;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Regulatory protein E2.
GN E2.
OS Human papillomavirus type 61.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=37116;
RN [1]
RP SEQUENCE FROM N.A.
RA Delius H.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.

CC -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
 CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACNNNNNNNGGT-3') PRESENT
 CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
 CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
 CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
 CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
 CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
 CC REPLICATION.
 CC -!- SUBUNIT: Binds DNA as a dimer.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U31793; AAA79495.1; -.
 DR HSSP; P17383; 1DHM.
 DR InterPro; IPR000427; E2_C.
 DR InterPro; IPR001866; E2_N.
 DR Pfam; PF00511; E2_C; 1.
 DR Pfam; PF00508; E2_N; 1.
 DR ProDom; PD000672; E2_C; 1.
 DR ProDom; PD000678; E2_N; 1.
 KW Early protein; Transcription regulation; Activator; DNA-binding;
 KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
 SQ SEQUENCE 382 AA; 43944 MW; 417F441DD7B772B4 CRC64;

Query Match 67.3%; Score 33; DB 1; Length 382;
 Best Local Similarity 75.0%; Pred. No. 27;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LSSRLDAC 9
 |: |||||
 Db 6 LADRLDAC 13

RESULT 14
 VE2_HPV32
 ID VE2_HPV32 STANDARD; PRT; 394 AA.
 AC P36791;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Regulatory protein E2.
 GN E2.
 OS Human papillomavirus type 32.
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
 OC Papillomavirus.
 OX NCBI_TaxID=10612;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94265501; PubMed=8205838;
 RA Delius H., Hofmann B.;

RT "Primer-directed sequencing of human papillomavirus types.";
 RL Curr. Top. Microbiol. Immunol. 186:13-31(1994).
 CC -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
 CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACNNNNNNNGGT-3') PRESENT
 CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
 CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
 CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
 CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
 CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
 CC REPLICATION.
 CC -!- SUBUNIT: Binds DNA as a dimer.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X74475; CAA52552.1; -.
 DR PIR; S36512; S36512.
 DR HSSP; P17383; 1DHM.
 DR InterPro; IPR000427; E2_C.
 DR InterPro; IPR001866; E2_N.
 DR Pfam; PF00511; E2_C; 1.
 DR Pfam; PF00508; E2_N; 1.
 DR ProDom; PD000672; E2_C; 1.
 DR ProDom; PD000678; E2_N; 1.
 KW Early protein; Transcription regulation; Activator; DNA-binding;
 KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
 SQ SEQUENCE 394 AA; 45038 MW; 113C46119C2265E7 CRC64;

Query Match 67.3%; Score 33; DB 1; Length 394;
 Best Local Similarity 75.0%; Pred. No. 28;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LSSRLDAC 9
 |: |||||
 Db 4 LAKRLDAC 11

RESULT 15

VE2_HPV42

ID VE2_HPV42 STANDARD; PRT; 398 AA.
 AC P27223;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Regulatory protein E2.
 GN E2.
 OS Human papillomavirus type 42.
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
 OC Papillomavirus.
 OX NCBI_TaxID=10590;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=92087479; PubMed=1309278;
 RA Philipp W., Honore N., Sapp M., Cole S.T., Streeck R.E.;
 RT "Human papillomavirus type 42: new sequences, conserved genome
 RT organization.";
 RL Virology 186:331-334(1992).
 CC -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
 CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACCNNNNNNGGT-3') PRESENT
 CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
 CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
 CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
 CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
 CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
 CC REPLICATION.
 CC -!- SUBUNIT: Binds DNA as a dimer.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M73236; AAA47044.1; ALT_INIT.
 DR PIR; B39451; W2WL42.
 DR HSSP; P17383; 1DHM.
 DR InterPro; IPR000427; E2_C.
 DR InterPro; IPR001866; E2_N.
 DR Pfam; PF00511; E2_C; 1.
 DR Pfam; PF00508; E2_N; 1.
 DR ProDom; PD000672; E2_C; 1.
 DR ProDom; PD000678; E2_N; 1.
 KW Early protein; Transcription regulation; Activator; DNA-binding;
 KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
 SQ SEQUENCE 398 AA; 45309 MW; 4D41D7196372808C CRC64;

Query Match 67.3%; Score 33; DB 1; Length 398;
 Best Local Similarity 75.0%; Pred. No. 28;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LSSRLDAC 9
 | : | | | |
 Db 4 LAKRLDAC 11

Search completed: November 13, 2003, 09:46:32
 Job time : 6.15625 secs

GenCore version 5.1.6
 Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 23.7188 Seconds

(without alignments)
97.917 Million cell updates/sec

Title: US-09-228-866-3
Perfect score: 49
Sequence: 1 CLSSRLDAC 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertibrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	% Query						Description
	No.	Score	Match	Length	DB	ID	
1	41	83.7	935	5	Q94719	Q94719	paramecium
2	41	83.7	2533	5	P90589	P90589	paramecium
3	41	83.7	2533	5	Q27183	Q27183	paramecium
4	41	83.7	2543	5	P90649	P90649	paramecium
5	37	75.5	92	16	Q9PDV6	Q9pdv6	xylella fas
6	37	75.5	152	16	Q9K6K3	Q9k6k3	bacillus ha
7	35	71.4	115	4	Q8TCB4	Q8tcb4	homo sapien
8	35	71.4	153	16	Q930G5	Q930g5	rhizobium m

9	35	71.4	161	5	Q9XV60	Q9xv60 caenorhabdi
10	35	71.4	251	12	Q8B3U7	Q8b3u7 porcine lym
11	35	71.4	258	17	Q9YC60	Q9yc60 aeropyrum p
12	35	71.4	389	16	Q9AAC8	Q9aac8 caulobacter
13	34	69.4	157	10	Q9LE05	Q9le05 medicago sa
14	34	69.4	185	6	O62685	O62685 saimiri sci
15	34	69.4	251	12	Q8JYB7	Q8jyb7 porcine lym
16	34	69.4	375	12	Q9DIH4	Q9dih4 human papil
17	34	69.4	494	16	Q8ZET7	Q8zet7 yersinia pe
18	34	69.4	529	16	Q8D0F3	Q8d0f3 yersinia pe
19	34	69.4	549	11	Q8BYV0	Q8byv0 mus musculu
20	34	69.4	781	3	Q99131	Q99131 ustilago ma
21	34	69.4	913	13	Q8AY18	Q8ay18 rana escule
22	34	69.4	1159	11	Q8CIS0	Q8cis0 mus musculu
23	34	69.4	1171	4	Q8TES3	Q8tes3 homo sapien
24	34	69.4	1650	5	Q8I2T7	Q8i2t7 plasmodium
25	34	69.4	1759	5	Q9XTP8	Q9xtp8 plasmodium
26	34	69.4	3306	10	Q9FT44	Q9ft44 arabidopsis
27	33.5	68.4	700	11	Q9DBD0	Q9dbd0 mus musculu
28	33.5	68.4	700	11	Q8VC96	Q8vc96 mus musculu
29	33	67.3	115	16	Q9RIZ0	Q9riz0 streptomyce
30	33	67.3	140	2	Q47504	Q47504 escherichia
31	33	67.3	162	5	Q9UA34	Q9ua34 ostrinia nu
32	33	67.3	162	5	Q9UA33	Q9ua33 ostrinia nu
33	33	67.3	162	5	Q9Y1I1	Q9y1i1 ostrinia fu
34	33	67.3	162	5	Q9TW42	Q9tw42 ostrinia nu
35	33	67.3	162	5	Q9TVH3	Q9tvh3 ostrinia nu
36	33	67.3	162	5	Q9Y1H9	Q9y1h9 ostrinia nu
37	33	67.3	162	5	Q9TW57	Q9tw57 ostrinia nu
38	33	67.3	162	5	Q9TVG5	Q9tvvg5 ostrinia nu
39	33	67.3	162	5	Q9Y1I0	Q9y1i0 ostrinia fu
40	33	67.3	197	10	Q8RV58	Q8rv58 oryza sativ
41	33	67.3	264	12	Q9DW32	Q9dw32 rat cytomeg
42	33	67.3	280	17	Q9YCP1	Q9ycp1 aeropyrum p
43	33	67.3	289	5	Q8IG38	Q8ig38 caenorhabdi
44	33	67.3	306	5	Q18896	Q18896 caenorhabdi
45	33	67.3	311	11	Q8R335	Q8r335 mus musculu

ALIGNMENTS

RESULT 1

Q94719

ID Q94719 PRELIMINARY; PRT; 935 AA.

AC Q94719;

DT 01-FEB-1997 (TrEMBLrel. 02, Created)

DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Epsilon 51-D i-ag (Fragment).

GN EPSILON-51D.

OS Paramecium tetraurelia.

OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;

OC Paramecium.

OX NCBI_TaxID=5888;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=51;
 RA Schwegmann K., Schulte G., Schmidt H.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; X96557; CAA65393.1; -.
 DR InterPro; IPR002895; Paramecium_SA.
 DR Pfam; PF01508; Paramecium_SA; 10.
 DR SMART; SM00639; PSA; 11.
 FT NON_TER 935 935
 SQ SEQUENCE 935 AA; 97925 MW; 2D75D09F28B44903 CRC64;

Query Match 83.7%; Score 41; DB 5; Length 935;
 Best Local Similarity 66.7%; Pred. No. 4.5;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
 |:|:|:|
 Db 575 CISNRVDAC 583

RESULT 2

P90589

ID P90589 PRELIMINARY; PRT; 2533 AA.
 AC P90589;
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Alpha-51D immobilization antigen.
 GN ALPHA-51D.
 OS Paramecium tetraurelia.
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
 OC Paramecium.
 OX NCBI_TaxID=5888;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=51;
 RA Schwegmann K., Klein H., Schmidt H.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; X96400; CAA65264.1; -.
 DR InterPro; IPR002895; Paramecium_SA.
 DR Pfam; PF01508; Paramecium_SA; 21.
 DR SMART; SM00639; PSA; 26.
 SQ SEQUENCE 2533 AA; 264142 MW; EAED7F21E408C371 CRC64;

Query Match 83.7%; Score 41; DB 5; Length 2533;
 Best Local Similarity 66.7%; Pred. No. 12;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
 |:|:|:|
 Db 575 CISNRVDAC 583

RESULT 3

Q27183

ID Q27183 PRELIMINARY; PRT; 2533 AA.
 AC Q27183;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)

QY 1 CLSSRLDAC 9
|:|:|:|
Db 575 CISNRVDAC 583

RESULT 5

Q9PDV6

ID Q9PDV6 PRELIMINARY; PRT; 92 AA.
AC Q9PDV6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein Xf1273.
GN Xf1273.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5c;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL; AE003961; AAF84082.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 92 AA; 9993 MW; 763E024E2E909A42 CRC64;

Query Match 75.5%; Score 37; DB 16; Length 92;
Best Local Similarity 87.5%; Pred. No. 3.3;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLSSRLDA 8
||:|||||
Db 67 CLASRLDA 74

RESULT 6

Q9K6K3

ID Q9K6K3 PRELIMINARY; PRT; 152 AA.
AC Q9K6K3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein BH3726.
GN BH3726.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001519; BAB07445.1; -.
DR InterPro; IPR000086; NUDIX_hydrolase.
DR Pfam; PF00293; NUDIX; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 152 AA; 17688 MW; 2C3A461F9CEDA6DD CRC64;

Query Match 75.5%; Score 37; DB 16; Length 152;
Best Local Similarity 77.8%; Pred. No. 5.4;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
|| ||| ||
Db 127 CLPSRLKAC 135

RESULT 7

Q8TCB4

ID Q8TCB4 PRELIMINARY; PRT; 115 AA.
AC Q8TCB4;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Strausberg R.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; BC022404; AAH22404.1; -.
 KW Hypothetical protein.
 FT NON_TER 1 1
 SQ SEQUENCE 115 AA; 12461 MW; 4E1BDD511025F61A CRC64;

Query Match 71.4%; Score 35; DB 4; Length 115;
 Best Local Similarity 66.7%; Pred. No. 11;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
 ||| |:
 Db 104 CLSQALEAC 112

RESULT 8

Q930G5

ID Q930G5 PRELIMINARY; PRT; 153 AA.
 AC Q930G5;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Hypothetical protein RA0231.
 GN RA0231 OR SMA0443.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OG Plasmid pSymA (megaplasmid 1).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396509; PubMed=11481432;
 RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
 RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
 RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
 RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
 RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
 RT "Nucleotide sequence and predicted functions of the entire
 RT Sinorhizobium meliloti pSymA megaplasmid."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
 DR EMBL; AE007216; AAK64889.1; -.
 KW Hypothetical protein; Plasmid; Complete proteome.
 SQ SEQUENCE 153 AA; 17025 MW; 0715B1AEF2C6FA04 CRC64;

Query Match 71.4%; Score 35; DB 16; Length 153;
 Best Local Similarity 77.8%; Pred. No. 14;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
 || |||
 Db 138 CLPSRLMAC 146

RESULT 9

Q9XV60

ID Q9XV60 PRELIMINARY; PRT; 161 AA.
 AC Q9XV60;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE F26D2.12 protein.
 GN F26D2.12.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA McMurray A.A.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL; Z81513; CAB04182.1; -.
 DR HSSP; P05140; 2AFP.
 DR WormPep; F26D2.12; CE18642.
 DR InterPro; IPR001304; Lectin_C.
 DR Pfam; PF00059; lectin_c; 1.
 DR SMART; SM00034; CLECT; 1.
 DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
 SQ SEQUENCE 161 AA; 16987 MW; B54C850333E98012 CRC64;

Query Match 71.4%; Score 35; DB 5; Length 161;
 Best Local Similarity 87.5%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLSSRLDA 8
 |||| |||
 Db 101 CLSSNLDA 108

RESULT 10

Q8B3U7

ID Q8B3U7 PRELIMINARY; PRT; 251 AA.
 AC Q8B3U7;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Uracil-DNA-glycosidase.
 OS Porcine lymphotropic herpesvirus 2.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae.
 OX NCBI_TaxID=91741;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=568;

RA Chmielewicz B., Goltz M., Franz T., Bauer C., Brema S., Ellerbrok H.,
 RA Beckmann S., Rziha H.-J., Lahrmann K.-H., Romero C., Ehlers B.;
 RT "A novel porcine gammaherpesvirus."
 RL Virology 0:0-0(2003).
 DR EMBL; AY170317; AA012387.1; -.
 KW Glycosidase.
 SQ SEQUENCE 251 AA; 28400 MW; F364582814E84F8B CRC64;

Query Match 71.4%; Score 35; DB 12; Length 251;
 Best Local Similarity 55.6%; Pred. No. 23;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
 |||::|::|
 Db 172 CLSNKLNLC 180

RESULT 11

Q9YC60

ID Q9YC60 PRELIMINARY; PRT; 258 AA.
 AC Q9YC60;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Hypothetical protein APE1391.
 GN APE1391.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
 OC Desulfurococcaceae; Aeropyrum.
 OX NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K1;
 RX MEDLINE=99310339; PubMed=10382966;
 RA Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix K1."
 RL DNA Res. 6:83-101(1999).
 DR EMBL; AP000061; BAA80388.1; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 258 AA; 28713 MW; 25C2FDB73F178AEB CRC64;

Query Match 71.4%; Score 35; DB 17; Length 258;
 Best Local Similarity 66.7%; Pred. No. 24;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
 ||| || |
 Db 223 CLSGRLSTC 231

RESULT 12

Q9AAC8

ID Q9AAC8 PRELIMINARY; PRT; 389 AA.
AC Q9AAC8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein CC0674.
GN CC0674.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of *Caulobacter crescentus*.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005743; AAK22659.1; -.
DR TIGR; CC0674; -.
DR InterPro; IPR000734; Lipase.
DR InterPro; IPR000897; SRP54.
DR PROSITE; PS00120; LIPASE_SER; 1.
DR PROSITE; PS00300; SRP54; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 389 AA; 40769 MW; C6DD05B8CE8D150E CRC64;

Query Match 71.4%; Score 35; DB 16; Length 389;
Best Local Similarity 66.7%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
|| | |||
Db 44 CLPGRADAC 52

RESULT 13

Q9LE05

ID Q9LE05 PRELIMINARY; PRT; 157 AA.
AC Q9LE05;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Resistance gene analog protein (Fragment).
OS Medicago sativa (Alfalfa).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
OX NCBI_TaxID=3879;
RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=cv. UC123;
 RA Cordero J.C., Skinner D.Z.;
 RT "Isolation of the Nucleotide Binding Site Family of Resistance Gene
 RT Analogs from Alfalfa, Medicago Sativa.";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AF230816; AAF72899.1; -.
 DR EMBL; AF230814; AAF72897.1; -.
 DR InterPro; IPR002182; NB-ARC.
 DR Pfam; PF00931; NB-ARC; 1.
 FT NON_TER 1 1
 FT NON_TER 157 157
 SQ SEQUENCE 157 AA; 17959 MW; 314D07D288F5A2A2 CRC64;

Query Match 69.4%; Score 34; DB 10; Length 157;
 Best Local Similarity 75.0%; Pred. No. 24;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LSSRLDAC 9
 :||| |||
 Db 9 ISSRFDAC 16

RESULT 14

O62685

ID O62685 PRELIMINARY; PRT; 185 AA.
 AC O62685;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE CD46 protein (Fragment).
 OS Saimiri sciureus (Common squirrel monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Saimiri.
 OX NCBI_TaxID=9521;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98184523; PubMed=9525611;
 RA Hsu E.C., Sarangi F., Iorio C., Sidhu M.S., Udem S.A., Dillehay D.L.,
 RA Xu W., Rota P.A., Bellini W.J., Richardson C.D.;
 RT "A single amino acid change in the hemagglutinin protein of measles
 RT virus determines its ability to bind CD46 and reveals another receptor
 RT on marmoset B cells.";
 RL J. Virol. 72:2905-2916(1998).
 DR EMBL; AF025483; AAC39671.1; -.
 DR HSSP; P10998; 1VVD.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00084; sushi; 2.
 DR SMART; SM00032; CCP; 2.
 FT NON_TER 185 185
 SQ SEQUENCE 185 AA; 20733 MW; A7CF613A0DF23742 CRC64;

Query Match 69.4%; Score 34; DB 6; Length 185;
 Best Local Similarity 87.5%; Pred. No. 28;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LSSRLDAC 9

Db

||||| |||
28 LSSRSDAC 35

RESULT 15

Q8JYB7

ID Q8JYB7 PRELIMINARY; PRT; 251 AA.
AC Q8JYB7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Uracil-DNA-glycosidase.
OS Porcine lymphotropic herpesvirus 1.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae.
OX NCBI_TaxID=91740;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=sample #56;
RX MEDLINE=99226949; PubMed=10211967;
RA Ehlers B., Ulrich S., Goltz M.;
RT "Detection of two novel porcine herpesviruses with high similarity to
RT gammaherpesviruses.";
RL J. Gen. Virol. 80:971-978(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=sample #56;
RX MEDLINE=20036635; PubMed=10567652;
RA Ulrich S., Goltz M., Ehlers B.;
RT "Characterization of the DNA polymerase loci of the novel porcine
RT lymphotropic herpesviruses 1 and 2 in domestic and feral pigs.";
RL J. Gen. Virol. 80:3199-3205(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=sample #56;
RX MEDLINE=22008485; PubMed=12009880;
RA Goltz M., Ericsson T., Patience C., Huang C.A., Noack S., Sachs D.H.,
RA Ehlers B.;
RT "Sequence analysis of the genome of porcine lymphotropic herpesvirus 1
RT and gene expression during post-transplant lymphoproliferative disease
RT of pigs.";
RL Virology 294:383-393(2002).
DR EMBL; AF478169; AAM22145.1; -.
DR InterPro; IPR002043; UDNA_glycsylse.
DR InterPro; IPR005122; UDNA_glycsylseSF.
DR InterPro; IPR003249; U_glycsylse_notp.
DR Pfam; PF03167; UDG; 1.
DR ProDom; PD001589; U_glycsylse_notp; 1.
DR TIGRFAMs; TIGR00628; ung; 1.
DR PROSITE; PS00130; U_DNA_GLYCOSYLASE; 1.
KW Glycosidase.
SQ SEQUENCE 251 AA; 28520 MW; 860234AE91746565 CRC64;

Query Match 69.4%; Score 34; DB 12; Length 251;
Best Local Similarity 55.6%; Pred. No. 38;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
 ||| :|::|
Db 172 CLSDKLNSC 180

Search completed: November 13, 2003, 09:50:59
Job time : 25.7188 secs

OM protein - protein search, using sw model

Run on: November 13, 2003, 10:27:36 ; Search time 10.125 Seconds
(without alignments)
37.610 Million cell updates/sec

Title: US-09-228-866-3
Perfect score: 9
Sequence: 1 CLSSRLDAC 9

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 118358

Minimum DB seq length: 7
Maximum DB seq length: 21

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query Match	Length			
1	9	100.0	9	1	US-08-526-710-3	Sequence 3, Appli
2	9	100.0	9	3	US-08-862-855-3	Sequence 3, Appli
3	9	100.0	9	3	US-09-226-985-3	Sequence 3, Appli
4	9	100.0	9	4	US-09-227-906-3	Sequence 3, Appli
5	7	77.8	7	1	US-08-526-710-19	Sequence 19, Appl
6	7	77.8	7	3	US-08-862-855-19	Sequence 19, Appl
7	7	77.8	7	3	US-09-226-985-19	Sequence 19, Appl
8	7	77.8	7	4	US-09-227-906-19	Sequence 19, Appl
9	5	55.6	17	4	US-08-469-260A-241	Sequence 241, App
10	5	55.6	17	4	US-08-488-446-241	Sequence 241, App
11	5	55.6	17	4	US-08-467-344A-241	Sequence 241, App

12	4	44.4	7	1	US-08-503-062-14	Sequence 14, Appl
13	4	44.4	7	3	US-09-461-697-330	Sequence 330, App
14	4	44.4	7	5	PCT-US96-11495-14	Sequence 14, Appl
15	4	44.4	8	1	US-07-834-848-14	Sequence 14, Appl
16	4	44.4	8	1	US-08-526-710-12	Sequence 12, Appl
17	4	44.4	8	2	US-09-016-366A-32	Sequence 32, Appl
18	4	44.4	8	2	US-09-016-366A-63	Sequence 63, Appl
19	4	44.4	8	2	US-08-978-404B-54	Sequence 54, Appl
20	4	44.4	8	2	US-08-687-219B-3	Sequence 3, Appli
21	4	44.4	8	2	US-08-687-219B-4	Sequence 4, Appli
22	4	44.4	8	2	US-08-687-219B-5	Sequence 5, Appli
23	4	44.4	8	2	US-08-687-219B-6	Sequence 6, Appli
24	4	44.4	8	2	US-08-687-219B-7	Sequence 7, Appli
25	4	44.4	8	2	US-08-687-219B-19	Sequence 19, Appl
26	4	44.4	8	2	US-08-687-219B-20	Sequence 20, Appl
27	4	44.4	8	2	US-08-687-219B-21	Sequence 21, Appl
28	4	44.4	8	2	US-08-687-219B-22	Sequence 22, Appl
29	4	44.4	8	2	US-08-687-219B-23	Sequence 23, Appl
30	4	44.4	8	2	US-08-687-219B-35	Sequence 35, Appl
31	4	44.4	8	3	US-08-862-855-12	Sequence 12, Appl
32	4	44.4	8	3	US-09-226-985-12	Sequence 12, Appl
33	4	44.4	8	4	US-09-227-906-12	Sequence 12, Appl
34	4	44.4	9	2	US-08-286-861-19	Sequence 19, Appl
35	4	44.4	9	2	US-08-286-861-20	Sequence 20, Appl
36	4	44.4	10	2	US-08-792-553-8	Sequence 8, Appli
37	4	44.4	10	4	US-09-129-192C-42	Sequence 42, Appl
38	4	44.4	11	1	US-08-482-106-17	Sequence 17, Appl
39	4	44.4	11	1	US-08-482-107B-4	Sequence 4, Appli
40	4	44.4	11	2	US-08-823-932-4	Sequence 4, Appli
41	4	44.4	11	2	US-08-286-861-35	Sequence 35, Appl
42	4	44.4	11	3	US-09-064-703-15	Sequence 15, Appl
43	4	44.4	11	3	US-09-106-216-54	Sequence 54, Appl
44	4	44.4	11	3	US-09-171-590-1	Sequence 1, Appli
45	4	44.4	11	3	US-09-171-590-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-08-526-710-3

; Sequence 3, Application US/08526710

; Patent No. 5622699

; GENERAL INFORMATION:

; APPLICANT: Ruoslahti, Erkki

; APPLICANT: Pasqualini, Renata

; TITLE OF INVENTION: Method of Identifying Molecules That

; TITLE OF INVENTION: Home to a Selected Organ In Vivo

; NUMBER OF SEQUENCES: 44

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell and Flores

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego

; STATE: California

; COUNTRY: United States

; ZIP: 92122

; COMPUTER READABLE FORM:


```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,710
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: circular
; MOLECULE TYPE: peptide
US-08-526-710-3

```

```

Query Match          100.0%; Score 9; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      9; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 CLSSRLDAC 9
        |||||
Db      1 CLSSRLDAC 9

```

RESULT 2

US-08-862-855-3

```

; Sequence 3, Application US/08862855
; Patent No. 6068829
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/862,855
; FILING DATE:

```

```

; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2621
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: circular
; MOLECULE TYPE: peptide
US-08-862-855-3

```

```

Query Match          100.0%; Score 9; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 CLSSRLDAC 9
        |||||
Db      1 CLSSRLDAC 9

```

```

RESULT 3
US-09-226-985-3
; Sequence 3, Application US/09226985
; Patent No. 6296832
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Molecules That Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/226,985
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: circular
; MOLECULE TYPE: peptide
US-09-226-985-3

```

```

Query Match          100.0%; Score 9; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      9; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 CLSSRLDAC 9
        |||||
Db      1 CLSSRLDAC 9

```

RESULT 4

US-09-227-906-3

```

; Sequence 3, Application US/09227906
; Patent No. 6306365
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/227,906
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3424
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: circular
; MOLECULE TYPE: peptide
US-09-227-906-3

```

```

Query Match          100.0%; Score 9; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      9; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 CLSSRLDAC 9
        |||||
Db      1 CLSSRLDAC 9

```

RESULT 5

US-08-526-710-19

```

; Sequence 19, Application US/08526710
; Patent No. 5622699
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/526,710
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-526-710-19

```

```

Query Match          77.8%; Score 7; DB 1; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2 LSSRLDA 8
        |||||
Db      1 LSSRLDA 7

```

RESULT 6

US-08-862-855-19

```

; Sequence 19, Application US/08862855
; Patent No. 6068829
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/862,855
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2621
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-862-855-19

```

```

Query Match          77.8%; Score 7; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      2 LSSRLDA 8
        |||||
Db      1 LSSRLDA 7

```

RESULT 7

US-09-226-985-19

```

; Sequence 19, Application US/09226985
; Patent No. 6296832
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Molecules That Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/226,985
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-226-985-19

```

```

Query Match          77.8%; Score 7; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2 LSSRLDA 8
        |||||
Db      1 LSSRLDA 7

```

RESULT 8

```

US-09-227-906-19
; Sequence 19, Application US/09227906
; Patent No. 6306365
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/227,906
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3424
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-227-906-19

```

```

Query Match          77.8%; Score 7; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      7; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      2 LSSRLDA 8
        |||||
Db      1 LSSRLDA 7

```

RESULT 9

US-08-469-260A-241

; Sequence 241, Application US/08469260A

; Patent No. 6451578

; GENERAL INFORMATION:

; APPLICANT: JOHN N. SIMONS

; APPLICANT: TAMI J. PILOT-MATIAS

; APPLICANT: GEORGE J. DAWSON

; APPLICANT: GEORGE G. SCHLAUDER

; APPLICANT: SURESH M. DESAI

; APPLICANT: THOMAS P. LEARY

; APPLICANT: ANTHONY SCOTT MUERHOFF

; APPLICANT: JAMES C. ERKER

; APPLICANT: SHERI L. BUIJK

; APPLICANT: ISA K. MUSHAHWAR

; TITLE OF INVENTION: NON-A, NON-B. NON-C, NON-D, NON-E HEPATITIS

; TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE

; NUMBER OF SEQUENCES: 716

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D

; STREET: 100 ABBOTT PARK ROAD

; CITY: ABBOTT PARK

; STATE: IL

; COUNTRY: USA

; ZIP: 60064-3500

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible


```

;   OPERATING SYSTEM:  PC-DOS/MS-DOS
;   SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/469,260A
;     FILING DATE:
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/424,550
;     FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  POREMBSKI, PRISCILLA E.
;     REGISTRATION NUMBER:  33,207
;     REFERENCE/DOCKET NUMBER:  5527.PC.01
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  708-937-6365
;     TELEFAX:  708-938-2623
;   INFORMATION FOR SEQ ID NO:  241:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  17 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-469-260A-241

```

```

Query Match          55.6%;  Score 5;  DB 4;  Length 17;
Best Local Similarity 100.0%;  Pred. No. 20;
Matches      5;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

```

```

Qy      2 LSSRL 6
        |||||
Db      7 LSSRL 11

```

RESULT 10

```

US-08-488-446-241
; Sequence 241, Application US/08488446
; Patent No. 6558898
; GENERAL INFORMATION:
;   APPLICANT:  JOHN N. SIMONS
;   APPLICANT:  TAMI J. PILOT-MATIAS
;   APPLICANT:  GEORGE J. DAWSON
;   APPLICANT:  GEORGE G. SCHLAUDER
;   APPLICANT:  SURESH M. DESAI
;   APPLICANT:  THOMAS P. LEARY
;   APPLICANT:  ANTHONY SCOTT MUERHOFF
;   APPLICANT:  JAMES C. ERKER
;   APPLICANT:  SHERI L. BUIJK
;   APPLICANT:  ISA K. MUSHAHWAR
;   TITLE OF INVENTION:  NON-A, NON-B. NON-C, NON-D, NON-E HEPATITIS
;   TITLE OF INVENTION:  REAGENTS AND METHODS FOR THEIR USE
;   NUMBER OF SEQUENCES:  716
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  ABBOTT LABORATORIES  D377/AP6D
;     STREET:  100 ABBOTT PARK ROAD
;     CITY:  ABBOTT PARK
;     STATE:  IL
;     COUNTRY:  USA

```

```

; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,446
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/424,550
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMBSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5527.PC.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 241:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-488-446-241

```

```

Query Match          55.6%; Score 5; DB 4; Length 17;
Best Local Similarity 100.0%; Pred. No. 20;
Matches      5; Conservative    0; Mismatches    0; Indels      0; Gaps      0;

```

```

Qy      2 LSSRL 6
        |||||
Db      7 LSSRL 11

```

RESULT 11

US-08-467-344A-241

```

; Sequence 241, Application US/08467344A
; Patent No. 6586568

```

GENERAL INFORMATION:

```

; APPLICANT: JOHN N. SIMONS
;           TAMI J. PILOT-MATIAS
;           GEORGE J. DAWSON
;           GEORGE G. SCHLAUDER
;           SURESH M. DESAI
;           THOMAS P. LEARY
;           ANTHONY SCOTT MUERHOFF
;           JAMES C. ERKER
;           SHERI L. BUIJK
;           ISA K. MUSHAHWAR

```

```

; TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
;                   REAGENTS AND METHODS FOR THEIR USE

```

```

; NUMBER OF SEQUENCES: 716

```

```

; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D

```

```

; STREET: 100 ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,344A
; FILING DATE: 07-Jun-1995
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/424,550
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMBSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5527.PC.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 241:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 241:
US-08-467-344A-241

```

```

Query Match          55.6%;  Score 5;  DB 4;  Length 17;
Best Local Similarity 100.0%;  Pred. No. 20;
Matches      5;  Conservative    0;  Mismatches    0;  Indels      0;  Gaps      0;

```

```

Qy      2 LSSRL 6
        |||||
Db      7 LSSRL 11

```

RESULT 12

US-08-503-062-14

; Sequence 14, Application US/08503062

; Patent No. 5723300

; GENERAL INFORMATION:

; APPLICANT: Denis, Gerald V.

; APPLICANT: Green, Michael R.

; TITLE OF INVENTION: NUCLEAR LOCALIZED TRANSCRIPTION FACTOR

; TITLE OF INVENTION: KINASES AND DIAGNOSTIC ASSAYS RELATED THERETO

; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

```

; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/503,062
; FILING DATE: 10-JUL-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 04020/080001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-503-062-14

```

```

Query Match          44.4%; Score 4; DB 1; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      4; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      2 LSSR 5
        ||||
Db      3 LSSR 6

```

RESULT 13

```

US-09-461-697-330
; Sequence 330, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSEQ for Windows Version 4.0

```

; SEQ ID NO 330
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-330

Query Match 44.4%; Score 4; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 SRLD 7
|||
Db 4 SRLD 7

RESULT 14

PCT-US96-11495-14

; Sequence 14, Application PC/TUS9611495
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF MASSACHUSETTS MEDICAL CENTER
; TITLE OF INVENTION: NUCLEAR LOCALIZED TRANSCRIPTION FACTOR
; TITLE OF INVENTION: KINASES AND DIAGNOSTIC ASSAYS RELATED THERETO
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/11495
; FILING DATE: 03-JUL-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/503,062
; FILING DATE: 10-JUL-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 04020/080W01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear

; MOLECULE TYPE: protein
PCT-US96-11495-14

Query Match 44.4%; Score 4; DB 5; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LSSR 5
|||
Db 3 LSSR 6

RESULT 15

US-07-834-848-14

; Sequence 14, Application US/07834848

; Patent No. 5436221

; GENERAL INFORMATION:

; APPLICANT: KITAGUCHI, HIROSHI

; APPLICANT: KOMAZAWA, HIROYUKI

; APPLICANT: KOJIMA, MASAYOSHI

; APPLICANT: MORI, HIDETO

; APPLICANT: NISHIKAWA, NAUYUKI

; APPLICANT: SATOH, HIDEAKI

; APPLICANT: ORIKASA, ATSUSHI

; APPLICANT: ONO, MITSUNORI

; APPLICANT: AZUMA, ICHIRO

; APPLICANT: SAIKI, IKUO

; TITLE OF INVENTION: PEPTIDE DERIVATIVES AND APPLICATION

; TITLE OF INVENTION: THEREOF

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak, & Seas

; STREET: 2100 Pennsylvania Ave., NW

; CITY: Washington

; STATE: DC

; COUNTRY: USA

; ZIP: 20037-3202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/834,848

; FILING DATE: 19920213

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Biggart, Waddell A.

; REGISTRATION NUMBER: 24,861

; REFERENCE/DOCKET NUMBER: Q28480

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)293-7060

; TELEFAX: (202)293-7860

; TELEX: 6491103

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 8 amino acids

; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-834-848-14

Query Match 44.4%; Score 4; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 SRLD 7
|||
Db 4 SRLD 7

Search completed: November 13, 2003, 10:41:54
Job time : 10.125 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:39:50 ; Search time 10.6875 Seconds
(without alignments)
35.630 Million cell updates/sec

Title: US-09-228-866-3
Perfect score: 49
Sequence: 1 CLSSRLDAC 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	49	100.0	9	1	US-08-526-710-3	Sequence 3, Appli
2	49	100.0	9	3	US-08-862-855-3	Sequence 3, Appli
3	49	100.0	9	3	US-09-226-985-3	Sequence 3, Appli
4	49	100.0	9	4	US-09-227-906-3	Sequence 3, Appli
5	38	77.6	130	4	US-08-936-165A-491	Sequence 491, App
6	37	75.5	357	4	US-09-252-991A-28480	Sequence 28480, A
7	35	71.4	232	4	US-09-252-991A-26162	Sequence 26162, A
8	35	71.4	3177	2	US-08-477-451-4	Sequence 4, Appli
9	34	69.4	781	1	US-08-373-134D-2	Sequence 2, Appli
10	34	69.4	781	2	US-09-114-637-2	Sequence 2, Appli
11	33	67.3	44	1	US-08-530-010-33	Sequence 33, Appli

12	33	67.3	44	2	US-08-484-101B-33	Sequence 33, Appl
13	33	67.3	44	3	US-08-714-524D-33	Sequence 33, Appl
14	33	67.3	309	3	US-08-969-644-20	Sequence 20, Appl
15	33	67.3	309	3	US-08-444-189-20	Sequence 20, Appl
16	33	67.3	309	3	US-08-468-544-20	Sequence 20, Appl
17	33	67.3	617	3	US-09-188-930-303	Sequence 303, App
18	33	67.3	617	4	US-09-312-283C-303	Sequence 303, App
19	32	65.3	143	4	US-09-461-325-203	Sequence 203, App
20	32	65.3	356	4	US-09-252-991A-27056	Sequence 27056, A
21	31.5	64.3	142	4	US-09-198-452A-807	Sequence 807, App
22	31	63.3	7	1	US-08-526-710-19	Sequence 19, Appl
23	31	63.3	7	3	US-08-862-855-19	Sequence 19, Appl
24	31	63.3	7	3	US-09-226-985-19	Sequence 19, Appl
25	31	63.3	7	4	US-09-227-906-19	Sequence 19, Appl
26	31	63.3	80	4	US-09-107-532A-4249	Sequence 4249, Ap
27	31	63.3	321	4	US-09-252-991A-18759	Sequence 18759, A
28	31	63.3	439	3	US-08-975-762-60	Sequence 60, Appl
29	31	63.3	439	3	US-09-295-028-60	Sequence 60, Appl
30	31	63.3	439	4	US-09-106-582-60	Sequence 60, Appl
31	31	63.3	677	4	US-09-252-991A-25571	Sequence 25571, A
32	31	63.3	1587	4	US-09-000-094-46	Sequence 46, Appl
33	31	63.3	1727	2	US-08-477-451-10	Sequence 10, Appl
34	31	63.3	4544	1	US-08-469-486-52	Sequence 52, Appl
35	31	63.3	4544	2	US-08-469-658-52	Sequence 52, Appl
36	30	61.2	143	4	US-09-252-991A-27539	Sequence 27539, A
37	30	61.2	181	4	US-09-252-991A-21966	Sequence 21966, A
38	30	61.2	217	6	5189147-7	Patent No. 5189147
39	30	61.2	264	1	US-08-448-606-6	Sequence 6, Appli
40	30	61.2	266	6	5175384-11	Patent No. 5175384
41	30	61.2	293	6	5189147-3	Patent No. 5189147
42	30	61.2	303	4	US-09-509-347-7	Sequence 7, Appli
43	30	61.2	306	3	US-09-082-593-2	Sequence 2, Appli
44	30	61.2	310	3	US-08-897-097-4	Sequence 4, Appli
45	30	61.2	311	3	US-08-897-097-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-08-526-710-3

; Sequence 3, Application US/08526710

; Patent No. 5622699

; GENERAL INFORMATION:

; APPLICANT: Ruoslahti, Erkki

; APPLICANT: Pasqualini, Renata

; TITLE OF INVENTION: Method of Identifying Molecules That

; TITLE OF INVENTION: Home to a Selected Organ In Vivo

; NUMBER OF SEQUENCES: 44

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell and Flores

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego

; STATE: California

; COUNTRY: United States

; ZIP: 92122

; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,710
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: circular
; MOLECULE TYPE: peptide
US-08-526-710-3

```

```

Query Match          100.0%; Score 49; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      9; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 CLSSRLDAC 9
        |||||
Db      1 CLSSRLDAC 9

```

RESULT 2

```

US-08-862-855-3
; Sequence 3, Application US/08862855
; Patent No. 6068829
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/862,855
; FILING DATE:

```

```

; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2621
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: circular
; MOLECULE TYPE: peptide
US-08-862-855-3

```

```

Query Match          100.0%; Score 49; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      9; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 CLSSRLDAC 9
        |||||
Db      1 CLSSRLDAC 9

```

```

RESULT 3
US-09-226-985-3
; Sequence 3, Application US/09226985
; Patent No. 6296832
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Molecules That Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/226,985
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: circular
; MOLECULE TYPE: peptide
US-09-226-985-3

```

```

Query Match          100.0%; Score 49; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      9; Conservative    0; Mismatches    0; Indels      0; Gaps      0;

```

```

Qy      1 CLSSRLDAC 9
        |||||
Db      1 CLSSRLDAC 9

```

RESULT 4

US-09-227-906-3

```

; Sequence 3, Application US/09227906
; Patent No. 6306365
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/227,906
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3424
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: circular
; MOLECULE TYPE: peptide
US-09-227-906-3

```

```

Query Match          100.0%; Score 49; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      9; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 CLSSRLDAC 9
        |||||
Db      1 CLSSRLDAC 9

```

RESULT 5

```

US-08-936-165A-491
; Sequence 491, Application US/08936165A
; Patent No. 6348582
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Burnham, Martin
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Lonetto, Michael
; APPLICANT: Nicholas, Richard
; APPLICANT: Pratt, Julie
; APPLICANT: Reichard, Richard
; APPLICANT: Rosenberg, Martin
; APPLICANT: Ward, Judith
; TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,
; TITLE OF INVENTION: Polypeptides and Their Uses
; NUMBER OF SEQUENCES: 534
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA

```

```

; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,165A
; FILING DATE: 24-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/027,032
; FILING DATE: 24-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50549
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 491:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 130 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-936-165A-491

```

```

Query Match          77.6%; Score 38; DB 4; Length 130;
Best Local Similarity 77.8%; Pred. No. 3.5;
Matches      7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      1 CLSSRLDAC 9
        || || ||
Db      112 CLLSRCDAC 120

```

```

RESULT 6
US-09-252-991A-28480
; Sequence 28480, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

```

; SEQ ID NO 28480
; LENGTH: 357
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28480

Query Match 75.5%; Score 37; DB 4; Length 357;
Best Local Similarity 87.5%; Pred. No. 16;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LSSRLDAC 9
:|||||||
Db 298 VSSRLDAC 305

RESULT 7

US-09-252-991A-26162
; Sequence 26162, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26162
; LENGTH: 232
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26162

Query Match 71.4%; Score 35; DB 4; Length 232;
Best Local Similarity 66.7%; Pred. No. 24;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
| ||: |||
Db 128 CRSSKADAC 136

RESULT 8

US-08-477-451-4
; Sequence 4, Application US/08477451
; Patent No. 5928865
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; TITLE OF INVENTION: Helicobacter Pylori Cagi Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation

; STREET: 4560 Horton Street
 ; CITY: Emeryville
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94608-2916
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/477,451
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: McClung, Barbara G.
 ; REGISTRATION NUMBER: 33,113
 ; REFERENCE/DOCKET NUMBER: 0335.002
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 510-601-2708
 ; TELEFAX: 510-655-3542
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3177 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-08-477-451-4

Query Match 71.4%; Score 35; DB 2; Length 3177;
 Best Local Similarity 77.8%; Pred. No. 3.7e+02;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9
 | | | | |
 Db 1497 CESSPLDAC 1505

RESULT 9

US-08-373-134D-2

; Sequence 2, Application US/08373134D
 ; Patent No. 5780296
 ; GENERAL INFORMATION:
 ; APPLICANT: Kmiec, Eric
 ; APPLICANT: Holloman, William
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS TO PROMOTE
 ; TITLE OF INVENTION: HOMOLOGOUS RECOMBINATION IN EUKARYOTIC CELLS AND
 ORGANISMS
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036-2711